

BY
(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number
WO 01/23541 A2

(51) International Patent Classification⁷: C12N 15/00 (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(21) International Application Number: PCT/US00/27065

(22) International Filing Date: 2 October 2000 (02.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/156,953 30 September 1999 (30.09.1999) US

(71) Applicant (for all designated States except US): ALEXION PHARMACEUTICALS, INC. [US/US]; Suite 360, 25 Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): FODOR, William, L. [US/US]; 236 Wildcat Road, Madison, CT 06443 (US). RAMSOONDAR, Jagdeece, J. [IT/US]; 46 Madison Street, Hamilton, NY 13346 (US).

(74) Agents: WIXON, Henry, N. et al.; Hale and Dorr LLP, 1455 Pennsylvania Avenue, N.W., Washington, DC 20004 (US).

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/23541 A2

BEU: AVAILABLE COPY

(54) Title: COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

(57) Abstract: Disclosed are novel compositions and methods useful for modulating expression of a target gene in a cell by insertion of exogenous DNA sequence into the target gene. The compositions and methods of the invention are useful for generation of knockout animals including mammals.

COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

TECHNICAL FIELD

The present invention is directed generally to the biological sciences, including 5 recombinant genetics and immunology. More particularly, there are described herein non-human knockout animals, preferably mammals, in which the expression of one or more genes has been altered. Also provided herein are xenograft transplants in which the expression of one or more genes has been modulated to prevent or reduce the likelihood of rejection by the transplant recipient.

10

BACKGROUND

The immune response of mammals, including humans, against invading pathogens, 15 toxins, and other foreign substances involves many specialized cells that act together. Lymphocytes are a class of white blood cells responsible for the specificity of the immune system. Two important classes of lymphocytes are T cells and B cells. T cells develop in the thymus, and are responsible for cell mediated immunity. There are many types of specialized T 20 cells, such as for example, helper T cells (which enhance the activity of other types of white blood cells), suppressor T cells (which suppress the activity of other white blood cells), and cytotoxic T cells (which kill cells). B cells develop in the bone marrow and exert their effect by producing and secreting antibodies.

A key to the coordinated immune response is complement, which, as described in U.S. Patent No. 5,679,345, is involved in the pathogenesis of tissue injury observed in many immunologically mediated diseases, such as systemic lupus, erythematosus, rheumatoid arthritis,

and immune-hemolytic anemia. Complement is also involved in rejection of transplanted organ grafts. Complement is responsible for much of the tissue injury in transplantation due to inflammatory conditions resulting from rejection or superimposed by infection, ischemia, and thrombosis of vessels in the graft, as well as tissue injury due to inflammation from similar 5 causes in patients who have not received an organ transplant. In particular, complement attack on cells is central to the rapid onset phase of immune mediated graft rejection (hyperacute rejection), where complement activation and subsequent tissue damage occur within hours.

Graft rejection may occur through a number of different mechanisms, with the time course of rejection being characteristic of the particular mechanism. Early rejection (hyperacute 10 rejection), occurring within minutes or hours of transplantation, involves complement activation by components that are present at the time of the transplant operation. Activation may occur via the classical pathway by preformed antibodies that are reactive with the "foreign" or non-self markers of the graft or via the alternative pathway in response to tissue damage in the graft as a result of, for example, ischemic damage to the organ during storage before transplantation.

15 Acute rejection occurs days to weeks after transplantation, and is caused by sensitization of the host to the foreign tissue that makes up the graft. Once the host's immune system has identified the transplanted tissue as foreign, all the resources of the immune system are marshaled against the graft, including both specific (antibody and T cell-dependent) responses and non-specific (phagocytic and complement-dependent) responses. Chronic rejection will usually only occur 20 when the graft recipient is immune-suppressed. Then the graft may survive long enough for tissue to undergo changes which ultimately affect survival of the graft. Such changes include hyperplasia and tissue hypertrophy, and endothelial cell damage leading to narrowing of the vascular lumen and potentially impairing the oxygen supply of the graft tissue.

Xenograft rejection of pig tissue is triggered by natural human antibodies that recognize carbohydrate xeno-antigens, such as Gal α (1,3) galactose, which is expressed on pig endothelial cells that line blood vessels. Weiss, *Science*, 285(20):1221-1222 (August 20, 1999). U.S. Patent No. 5,821,117 describes inhibiting xenotransplant rejection by disrupting the wild type porcine Gal α (1,3) galactosyl transferase gene with a cloned mutant porcine Gal α (1,3) galactosyl transferase sequence specifically within an exon of the wild type gene. The resultant mutant gene does not encode a functional galactosyl transferase, with the expected result that rejection of the transplanted xenograft by the patient's immune system is avoided.

In such so called "knockout" mammals, expression of an endogenous gene has been altered (typically, suppressed) through genetic manipulation. Preparation of knockout mammals typically has required introducing into an undifferentiated cell type (termed an embryonic stem cell) a nucleic acid construct to suppress expression of a target gene. This cell is introduced and integrated into a mammalian embryo. The embryo is implanted into a foster mother for the duration of gestation. For example, Pfeffer et al. (*Cell*, 73:457-467 [1993]) describe mice in which the gene encoding the tumor necrosis factor receptor p55 has been disrupted by mutation utilizing homologous recombination. The mice showed a decreased response to tumor necrosis factor signaling. Fung-Leung et al. (*Cell*, 65:443-449 [1991]; *J. Exp. Med.*, 174:1425-1429 [1991]) describe knockout mice lacking expression of the gene encoding CD8. These mice were found to have a decreased level of cytotoxic T cell response to various antigens and to certain viral pathogens such as lymphocytic choriomeningitis virus.

Typical prior methods, however, describe manipulation of an exon region of the target gene. There is thus a need in the art for new and improved methods for modulating gene

expression in animals including mammals, particularly for overcoming xenograft transplant rejection. It is to these, as well as other, important ends that the following is addressed.

SUMMARY

It has been discovered that the expression of a particular gene in an animal may be 5 modulated by introducing into the genomic DNA of the animal a new DNA sequence that results in the disruption of at least some portion of the DNA sequence of the gene to be modulated. The methods described herein are of general utility for altering gene expression in animals including mammals. In contrast to prior methods, it has surprisingly been found that gene expression may be suppressed in part or in total by inserting new DNA sequence into the intron of the target 10 genomic DNA.

The versatility of the methods described herein for generating "knockout" animals is illustrated by the following general description of a preferred embodiments, including the examples. It is to be understood that while the remaining discussion is directed largely to the utility of Gal α (1,3)galactosyl transferase knockout pigs, the utility of the methods described 15 herein is not limited to solely this protein. Rather, the following discussion is provided merely for exemplification of their versatility and preferred use.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the nucleotide sequence of introns of the porcine Gal α (1,3) galactosyl transferase gene from within intron 3 to the end of intron 8. Dashes indicate nucleotides within 20 an exon region. Thus, nucleotide sequence numbering represents the number of bases in the entire porcine Gal α (1,3) galactosyl transferase gene relative to nucleotide position 1 of the insert isolated from the lambda-2 phage clone.

FIG. 2 shows a schematic representation of the gene targeting vector used for inactivation of the porcine Gal α (1,3) galactosyl transferase gene (see Example 1). This vector is designed to contain a sequence with homology to the 5' region of intron 3 of the Gal α (1,3) galactosyl transferase gene, a promoterless neomycin phosphotransferase gene engineered to contain multiple 5 stop codons (engineered exon), an engineered splice acceptor site, the 5' region of intron 4 sequence for splicing the engineered exon to the downstream exon 4, and a sequence with homology to the 3' region of intron 3 to aid with annealing to the porcine Gal α (1,3) galactosyl transferase gene. Arrows indicate location of primers used for PCR.

FIG. 3 shows the nucleotide sequence of the gene targeting vector used for inactivation of the porcine Gal α (1,3) galactosyl transferase gene (see Example 1). This vector is designed to contain (A.) a sequence with homology to the 5' region of intron 3 of the Gal α (1,3) galactosyl transferase gene, (B.) an intron 4 splice acceptor sequence, (C.) a promoterless neomycin phosphotransferase gene engineered to contain multiple stop codons (engineered exon), (D.) an intron 4 splice donor signal sequence, and (E.) a 3' intron 3 sequence to aid with annealing to the porcine Gal α (1,3) galactosyl transferase gene. All underlined sequences correspond to restriction sites in the primer sequences. Bold type indicates primer regions used for PCR. Normal type indicates PCR fragment sequences.

FIG. 4 shows the nucleotide sequence for the neomycin phosphotransferase gene (the neomycin resistance gene). Bold type indicates the location of gene start and stop codons. The 20 underlined sequence corresponds to primer sequences. Nucleotides which are capitalized are within the coding region of this gene.

FIG. 5 shows the nucleotide sequence of the puromycin/bovine growth hormone poly A. The underlined sequences correspond to the puromycin gene start codon.

FIG. 6 shows a schematic representation of the gene targeting vector used for inactivation of the porcine Gal α (1,3) galactosyl transferase gene (see Example 2). This vector is designed to contain a sequence with homology to the Gal α (1,3) galactosyl transferase gene 3' intron 3 sequence including the 3' intron splice acceptor sequence, a Kozak consensus sequence, a 5 promoterless puromycin gene engineered to contain a bovine growth hormone poly A sequence (engineered exon), and a sequence with 5' intron 4 sequence homology including the 5' intron splice donor sequence. Arrows indicate location of primers used for PCR.

FIG. 7 shows the nucleotide sequence of the gene targeting vector shown schematically in FIG. 6 (see Example 2). The underlined sequences correspond to the primer sequences used. 10 Bold type indicates the intron regions used for homology. The AG and GT splice consensus sequences at the 3' end of intron 3 and the 5' end of intron 4 are in upper case.

FIG. 8 shows the nucleotide sequence of the ricin A toxin gene. Nucleotides which are capitalized are within the coding region of this gene.

FIG. 9 shows a schematic representation of the collision construct used for inactivation of 15 the porcine Gal α (1,3) galactosyl transferase gene (see Example 3). This vector is designed to contain a sequence with homology to the Gal α (1,3) galactosyl transferase gene 3' intron 3 sequence including the 3' intron splice acceptor sequence, a reverse orientation puromycin gene engineered to contain a bovine growth hormone poly A sequence under the control of a phosphoglycerate kinase (PGK) promoter, and a sequence with 5' intron 4 sequence homology 20 including the 5' intron splice donor sequence, and a ricin A toxin gene under the control of a cytomegalovirus (CMV) promoter and containing a SV40 poly A sequence located outside the regions of homology. Arrows indicate location of primers used for PCR.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art, unless otherwise defined herein. Although any methods and materials similar or equivalent to those described herein may be used in the practice or testing of the described methods, preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection herewith.

As used herein and in the appended claims, the singular forms "a," "an," and "the" are intended to include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" is intended to include a plurality of such host cells, reference to "an antibody" is intended as a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth. It is to be understood that the appended claims are not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, which those of skill will appreciate may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is to be limited only by the appended claims.

The term "knockout" refers to the modulation of the expression of at least a portion of a protein encoded by the target gene. The term "knockout construct" refers to a nucleic acid sequence that is designed to modulate a protein encoded by endogenous DNA sequences in a cell. The nucleic acid sequence used as the knockout construct is typically comprised of DNA from some portion of the gene or genes (including, but not limited to, the exon sequence, intron

sequence, and/or promoter sequence) to be modulated and a sequence marker used to disrupt and select for the presence of the knockout construct in the cell. The nucleic acid sequence of the knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt protein expression from the native gene. Such 5 insertion usually occurs by homologous recombination (i.e., regions of the knockout construct that are homologous to endogenous DNA sequences hybridize to each other when the knockout construct is introduced into the cell and recombines so that the knockout construct is incorporated into the corresponding position of the endogenous DNA).

The knockout construct nucleic acid sequence may comprise a full or partial sequence of 10 one or more exons and/or introns of the gene to be modulated, a full or partial promoter sequence of the gene to be modulated, or combinations thereof. In one embodiment of the invention, the nucleic acid sequence of the knockout construct comprises a first nucleic acid sequence region homologous to a first nucleic acid sequence region of the gene to be modulated, and a second nucleic acid sequence region homologous to a second nucleic acid sequence region of the gene to 15 be modulated. The orientation of the knockout construct should be such that the first nucleic acid sequence is upstream of the second nucleic acid sequence and the sequence marker should be therebetween.

A suitable nucleic acid sequence region(s) should be selected so that there is homology 20 between knockout construct sequence(s) and the gene of interest. Preferably, the knockout construct sequences are isogenic sequences with respect to the target sequences. The nucleic acid sequence region of the knockout construct may correlate to any region of the gene provided that it is homologous to the gene. A nucleic acid sequence is considered to be "homologous" if it is at least about 90% identical, preferably at least about 95% identical, or most preferably, about

100% identical to the nucleic acid sequence. Furthermore, the 5' and 3' nucleic acid sequences flanking the selectable marker should be sufficiently large to provide complementary sequence for hybridization when the knockout construct is introduced into the genomic DNA of the target cell. For example, homologous nucleic acid sequences flanking the selectable marker gene 5 should be at least about 500 bp, preferably, at least about 1 kilobase (kb), more preferably about 2-4 kb, and most preferably about 3-4 kb in length. In a preferred embodiment, both of the homologous nucleic acid sequences flanking the selectable marker gene of the construct should be should be at least about 500 bp, preferably, at least about 1 kb, more preferably about 2-4 kb, and most preferably about 3-4 kb in length.

10 Another suitable DNA sequence includes cDNA sequence provided the cDNA is sufficiently large. Each of the flanking nucleic acid sequences used to make the construct is preferably homologous to one or more exon and/or intron regions, and/or a promoter region. Each of these sequences is different from the other, but may be homologous to regions within the same exon and/or intron. Alternatively, these sequences may be homologous to regions within 15 different exons and/or introns of the gene. Preferably, the two flanking nucleic acid sequences of the knockout construct are homologous to two sequence regions of the same or different introns of the gene of interest. In addition, it is preferred that isogenic DNA is used to make the knockout construct of the present invention. Thus, the nucleic acid sequences obtained to make the knockout construct are preferably obtained from the same cell line as that being used as the 20 target cell.

In accordance with the present invention, the integration of the knockout construct nucleic acid sequence into at least one gene of interest results in the modulation of the expression of the gene product. "Modulating" the expression of a gene includes suppressing the expression

of the gene, disrupting the expression of the gene, eliminating the expression of the gene, altering the expression of the gene, or decreasing the expression of the gene relative to expression of the wild-type gene. Preferably, the integrated knockout construct results in reduced protein function relative to native protein function. Most preferably, the integrated knockout construct results in 5 the production of a non-functional protein. Complete or absolute non-functionality of the protein is not required.

The phrases "disruption of the gene" and "gene disruption" refer to insertion of a nucleic acid sequence into at least one region of the native DNA sequence (usually one or more exons or one or more introns) and/or the promoter region of a gene so as to modulate expression of that 10 gene in the cell as compared to the wild-type or naturally occurring sequence of the gene. By way of example, a nucleic acid construct may be prepared containing a DNA sequence encoding an antibiotic resistance gene which is inserted between the DNA sequence complementary to the target gene DNA sequence (promoter and/or coding region) to be disrupted. When this nucleic acid construct is then transfected into a cell, the construct will integrate into the genomic DNA 15 either randomly or into the target gene by homologous recombination. It has been found that selection for drug resistant cells in the population of transfectants enhance the probability of obtaining a homologous gene knockout. Thus, many progeny of the cell will no longer express the gene, or will express it at a decreased level, as the DNA is now disrupted by the antibiotic resistance gene.

20 In some instances, such as, for example, where the methods described herein are used to produce cells, tissues or organs suitable for xenotransplant into humans, it may not be necessary to completely eliminate the production of functional protein. Rather, it will be satisfactory to reduce the production of functional protein only to a level that will, in conjunction with other

therapeutic regimens, prevent or reduce the patient's immune response and the likelihood of rejection. Thus, for example, a knockout achieved according to the methods described herein may preferably reduce the biological activity of the polypeptide normally encoded therefrom by at least about 70%, preferably at least about 80%, relative to the unmutated gene.

5 The knockout construct may be inserted into any suitable target cell for integration into its genomic DNA that may be maintained in culture. Suitable cells include, but are not limited to, fibroblast, epithelial cell, endothelial cell, transgenic embryonic fibroblast, embryonic stem cell, and primordial germ cell. In one embodiment, the knockout construct is inserted into an embryonic stem cell (ES cell) and is integrated into the ES cell genomic DNA. ES cells
10 comprising the integrated knockout construct are then injected into, and integrate with, a developing mouse embryo. In another embodiment, the knockout construct is inserted into a nuclear transfer donor cell. Suitable nuclear transfer donor cells include fibroblasts, epithelial cells, and cumulous cells. In this embodiment, the knockout construct is inserted into the nuclear transfer donor cell, and the donor cells comprising the knockout construct are fused with an
15 enucleated oocyte. The resultant fused oocyte is then transferred to a surrogate female. Furthermore, where the target cell is intended to be used to produce a knockout mammal, it is preferred that the target cell be derived from the same species as the knockout mammal to be generated. Thus, for example, pig embryonic stem cells or pig fibroblasts will usually be used for generation of knockout pigs.

20 The nucleic acid sequence of the knockout construct may be integrated into the genomic DNA of the host cell using any suitable method. In one preferred embodiment, integration is achieved by the process of homologous recombination. Homologous recombination has been described previously, for example, in Kucherlapati *et al.* (1984) Proc. Natl. Acad. Sci. USA

81:3153-3157; Kucherlapati *et al.* (1985) Mol. Cell. Bio. 5:714-720; Smithies *et al.* (1985) Nature 317:230- 234; Wake *et al.* (1985) Mol. Cell. Bio. 8:2080-2089; Ayares *et al.* (1985) Genetics 111:375-388; Ayares *et al.* (1986) Mol. Cell. Bio. 7:1656- 1662; Song *et al.* (1987) Proc. Natl. Acad. Sci. USA 84:6820-6824; Thomas *et al.* (1986) Cell 44:419-428; Thomas and 5 Capecchi (1987) Cell 51:503- 512; Nandi *et al.* (1988) Proc. Natl. Acad. Sci. USA 85:3845-3849; and Mansour *et al.* (1988) Nature 336:348-352, which are herein incorporated by reference. Furthermore, various aspects of using homologous recombination to create specific genetic mutations in embryonic stem cells and to transfer these mutations to the germline have been described. (Evans and Kaufman (1981) Nature 294:154-146; Doetschman *et al.*, (1987) 10 Nature 330:576-578; Thomas and Capecchi (1987) Cell 51:503-512; Thompson *et al.* (1989) Cell 56:316-321.) In homologous recombination, DNA fragments between two DNA molecules are exchanged during crossover at the site of the homologous nucleic acid sequences. Thus, crossover would occur between the knockout construct and eukaryotic gene at the site of homology within the 5' region of the first nucleic acid sequence of the construct (homologous to 15 the first nucleic acid region of the gene of interest). A second crossover event would occur in the 3' region of the construct homologous to the second nucleic acid region of the gene of interest. As a result, the sequence information between these two regions of the knockout construct would be inserted into the gene of interest in the host cell's genomic DNA.

The methods described herein may be used to produce a mammal in which one, two, or 20 more genes have been knocked out. Such mammals may be generated by repeating the procedures set forth herein for generating each knockout construct, or by breeding two mammals, each with a different single gene knocked out, to each other, and screening for those with the double knockout genotype.

The term "marker sequence" or "selectable marker" refers to a nucleic acid sequence that is used as part of the knockout construct to modulate the expression of the gene of interest, and as a means to identify those cells that have incorporated the knockout construct into the genome. The selectable marker may be any sequence that serves these purposes. For example, the

5 selectable marker may encode a protein that confers a detectable trait on the cell, such as an antibiotic resistance gene, or an assayable enzyme not typically found in the target cell. The selectable marker gene may be any nucleic acid sequence that is detectable and/or assayable, which is used to recover transformed cell lines. One having skill in the art will be capable of determining suitable selectable markers for use in the present invention. For example, suitable

10 selectable markers include, but are not limited to, β -lactamase (ampicillan resistance), kanamycin resistance, gentecin resistance, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. For example, the herpes simplex virus thymidine kinase (tk) (Wigler, M. et al. (1977) Cell 11:223-32) or adenine phosphoribosyltransferase (aprt) (Lowy, I. et al. (1980) Cell 22:817-23) genes, which may be

15 employed in tk or aprt cells, respectively, may be used as the selectable marker. Also, antimetabolite, antibiotic or herbicide resistance may be used as the basis for selection; for example, dihydrofolate reductase (dhfr), which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); neomycin phosphotransferase (npt), which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described and include, for example,

20 tryptophan synthetase (trpB), which allows cells to utilize indole in place of tryptophan, or histidinol dehydrogenase (hisD), which allows cells to utilize histidinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047- 51). Recently, the

use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase (GUS), and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

5 In the present invention, it is preferred that the selectable marker gene is an antibiotic resistance gene, such as the neomycin resistance gene or puromycin resistance gene.

Moreover, when the selectable marker encodes a protein, it may also contain a promoter that regulates its expression, or require expression from an endogenous promoter, preferably the target gene promoter. Thus, the selectable marker gene may be operably linked to its own 10 promoter or be promoterless. The selectable marker gene may be inserted into the knockout construct without its own promoter attached as it may be transcribed using the promoter of the gene to be suppressed. In addition, the marker gene may have a polyA signal sequence attached to the 3' end of the gene, which serves to terminate transcription of the gene and process the transcript with the addition of adenine residues at the 3' end to stabilize the mRNA.

15 In one embodiment a target gene (*e.g.*, Gal α (1,3) galactosyl transferase) is modulated by insertion of an engineered exon or active gene within an intron of the target gene. In this embodiment, the target gene is prevented from being translated by insertion of an in-frame, promoterless engineered exon (*e.g.*, an antibiotic resistance gene) that contains multiple stop codons within an intron of the target gene. Using this 'promoter-trap' strategy, the engineered exon is spliced 20 in frame upstream of the exon comprising the start codon. This results in the expression of the drug resistance gene prior to the gene of interest and concomitantly inhibits expression of the target gene due to the presence of multiple stop codons downstream of the drug resistance gene. As described

herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon.

Using the 'promoter trap' strategy, a gene targeting construct is designed which contains a sequence with homology to an intron sequence of the target gene (e.g., the intron 3 sequence of 5 the Gal α (1,3) galactosyl transferase gene), a downstream intron splice acceptor signal sequence comprising the AG dinucleotide splice acceptor site (e.g., the intron 4 splice acceptor signal sequence of the Gal α (1,3) galactosyl transferase gene), a promoterless selectable marker engineered exon (e.g., drug resistance gene) engineered to contain multiple stop codons, the intron splice donor signal sequence comprising the GT dinucleotide splice donor site (e.g., the intron 4 splice donor sequence of 10 the Gal α (1,3) galactosyl transferase gene) for splicing the engineered exon to the immediate downstream exon (e.g., exon 4 of the Gal α (1,3) galactosyl transferase gene), and additional sequence with homology to the intron sequence of the target gene (e.g., intron 3 sequence homology of the Gal α (1,3) galactosyl transferase gene) to aid with annealing to the target gene. It will be appreciated that the method may be used to target any intron within target gene of interest.

15 In another embodiment, the 'promoter trap' strategy is used to modulate target gene expression by replacing an endogenous exon with an in-frame, promoterless engineered exon (e.g., an antibiotic resistance gene). The engineered exon is spliced in frame and results in the expression of the drug resistance gene and concomitant inhibited expression of the full-length target gene.

20 This 'promoter trap' gene targeting construct may be designed to contain a sequence with homology to the target gene 3' intron sequence upstream of the start codon (e.g., the Gal α (1,3) galactosyl transferase gene 3' intron 3 sequence), the upstream intron splice acceptor sequence comprising the AG dinucleotide splice acceptor site (e.g., the intron 3 splice acceptor sequence), a

Kozak consensus sequence, a promoterless selectable marker gene containing *e.g.*, a poly A termination sequence (the engineered exon), a splice donor sequence comprising the GT dinucleotide splice donor site from a intron region downstream of the start codon (*e.g.*, the 5' intron 4 splice donor sequence), and a sequence with 5' sequence homology to the downstream intron (*e.g.*, 5' intron 4). It 5 will be appreciated that the method may be used to target any exon within the Gal α (1,3) galactosyl transferase gene or any other gene of interest. A representative construct useful for targeting the pig Gal α (1,3) galactosyl transferase gene was deposited with the American Type Culture Collection (ATCC) on 28 September 2000 with accession number _____ and is described herein in Example 2.

In yet another embodiment, the selectable marker may be inserted into the knockout 10 construct in a reverse orientation to the targeted gene. In this embodiment, a strong promoter is used with the selectable marker all in the reverse orientation, which drives transcription in the reverse direction and therefore, modulates the expression of the targeted gene. The target gene is modulated using a "collision construct" to insert an active gene in place of an exon and at least part of the flanking introns, including the splice donor and splice acceptor sites. The inserted gene, such as a 15 selectable marker gene, is under the control of a highly active promoter such as the phosphoglycerate kinase I (PGK) gene promoter, such that transcription of this gene causes the termination of transcription of the endogenous gene (Rosario *et al.*, (1996) Nat. Biotech.14:1592-1596). The selectable marker gene is further engineered to contain a transcription termination sequence. Insertion of the engineered gene may be made to replace any exon, within any intron, or portions thereof to 20 result in a truncated transcript which modulates the expression of a functional target gene product. It will be appreciated that this method may be used to target any intron or exon of interest of the target gene. Positive selection for transfected cells in which the construct has been integrated may be accomplished via expression of the selectable marker gene. As described herein, it will be appreciated

that any selection marker gene that confers survival of the targeted cells under appropriate selection conditions may be driven by the strong PGK promoter. Additionally, a toxin gene (e.g., Ricin A toxin) is preferably engineered into the collision construct inserted to eliminate random integration events. A representative collision construct useful for targeting the pig Gal α (1,3) 5 galactosyl transferase gene has been deposited with the ATCC on 28 September 2000 with accession number _____ and is described herein in Example 3.

The integrated selectable marker nucleic acid in the cell is capable of modulating the expression of the gene of interest. Expression of the selectable marker allows for selection of the cells which comprise the integrated sequence. Modulation of the expression of the gene of 10 interest is accomplished by disruption of the endogenous gene by an engineered exon in forward or reverse orientation with the endogenous gene.

The term "animal," as used herein, is intended to include any multicellular eukaryotic organism, preferred among which are mammals. When used in the context of a xenograft donor, the term "mammal" preferably includes, but is not limited to, pigs, sheep, goats, cows, deer, 15 rabbits, hamsters, rats, mice, horses, cats, dogs, and the like. Preferably, humans are excluded.

The term "progeny" refers to any and all future generations derived or descending from a particular mammal, i.e., a mammal containing a knockout construct inserted into its genomic DNA. Thus, progeny of any successive generation are included herein such that the progeny, the F1, F2, F3 generations, and so on indefinitely, are included in this definition.

20 The terms "immunomodulate" and "immunomodulation" refer to changes in the level of activity of any components of the immune system as compared to the average activity of that component for a particular species. Thus, as used herein, immunomodulation refers to an increase or a decrease in activity. Preferably, in accordance with the present invention, the

integration of the selectable marker into the gene of interest results in a decreased immune response, when the host cell is introduced to a patient. Immunomodulation may be detected by assaying the level of antibody reactivity, complement activity, B cells, any or all types of T cells, antigen presenting cells, and any other cells believed to be involved in immune function.

5 Additionally or alternatively, immunomodulation may be detected by evaluating the level of expression of particular genes believed to have a role in the immune system, the level of particular compounds such as cytokines (interleukins and the like) or other molecules that have a role in the immune system, and/or the level of particular enzymes, proteins, and the like that are involved in immune system functioning.

10 The target gene to be knocked out may be any gene, provided that at least some sequence information on the DNA to be disrupted is available to use in the preparation of both the knockout construct and the screening probes. It is not necessary that the entire genomic sequence of the target gene be known in order to use the methods described herein.

15 The target gene to be knocked out preferably will be a gene that is expressed in mature and/or immature T cells and/or B cells. It is a further preference that the target gene is expressed in target antigen presenting cell, target endothelial cell, target neuronal cell, or any target cell that may be attacked by the humoral or cellular immune system of the recipient. The target gene is further preferably involved, either directly or indirectly, in the activation pathway during inflammation or immunosuppression responses by the immune system, and does not result in lethality when knocked out. In accordance with the present invention, expression of target genes may advantageously be altered according to the methods described herein to produce xenotransplant cells, tissues and organs for use in humans, in order to reduce or prevent immune response and rejection by the patient.

Thus, in accordance with the present invention, any gene may be used provided that it can undergo homologous recombination and the expression of which may be modulated by insertion of the knockout construct of the present invention. Suitable genes include, but are not limited to, B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2 or VCAM-1, CD28, CD80, CD86, 5 CD154, major histocompatibility complex class I, β -2-microglobulin, invariant chain (Ii), caspase-1, caspase-3, and Gal α (1,3) galactosyl transferase gene. This list is not intended to be exhaustive. One having ordinary skill in the art would be capable of ascertaining suitable genes to be modulated. Preferably, the gene is implicated in the immunoresponse system of a patient. More preferably, the target gene is a porcine target gene selected from the group consisting of 10 CD 80, CD 86, B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2 or VCAM-1. A presently preferred porcine target gene is the Gal α (1,3) galactosyl transferase gene.

The DNA sequence to be used to knock out a selected gene may be obtained using methods well known in the art such as those described by Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). 15 Such methods include, for example, screening a genomic library with a cDNA probe encoding at least a portion of the same gene in order to obtain at least a portion of the genomic sequence. Alternatively, if a cDNA sequence is to be used in a knockout construct, the cDNA may be obtained by screening a cDNA library with oligonucleotide probes or antibodies (where the library is cloned into an expression vector). If a promoter sequence is to be used in the knockout 20 construct, synthetic DNA probes may be designed for screening a genomic library containing the promoter sequence. Another method for obtaining the DNA to be used in the knockout construct is to manufacture the DNA sequence synthetically, using a DNA synthesizer.

In another embodiment, porcine genomic DNA encoding the Gal α (1,3) galactosyl transferase gene is isolated from a lambda phage clone library. A pig genomic library is screened using a cDNA corresponding to exon 4 of the Gal α (1,3) galactosyl transferase gene. Phage are screened and unique clones, which contain exon 4 sequences are isolated using standard library screening methods. (Sambrook et al.) Clones obtained by this procedure contain inserts 15-40 kb in length. These clones, were designated pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2. Five vectors comprising unique, overlapping nucleotide sequences which span the entire the pig Gal α (1,3) galactosyl transferase gene from within intron 3 through intron 8 have been deposited with the ATCC: (1) a 1.6 kb insert within intron 3 of the extreme 5' end of the 18.275 kb lambda-2 phage clone, (2) a 6.7 kb HindIII fragment spanning intron 3 to intron 4 of the 18.275 kb lambda-2 phage clone, (3) a 4 kb HindIII fragment following the 6.7 kb fragment 2 of the 18.275 lambda-2 phage clone, (4) a 6 kb HindIII-SalI fragment at the 3' most portion of the 18.275 lambda-2 phage clone, and (5) a 13 kb fragment of the lambda-2 phage clone spanning exon 7 to exon 9. These five vectors were deposited with ATCC on 29 September 2000 with accession numbers _____, respectively. Subclones of the various inserts were used to generate the claimed intron sequences from within intron 3 to intron 8 as provided in Figure 1. These sequences may be used to determine regions of sequence homology in design of targeting constructs for modulation of the pig Gal α (1,3) galactosyl transferase gene.

The DNA sequence encoding the knockout construct is preferably generated in sufficient quantity for genetic manipulation and insertion the target cell. Amplification may be accomplished by known methods, such as by placing the sequence into a suitable vector and transforming bacterial or other cells that may rapidly amplify the vector, by PCR amplification, or by synthesis with a DNA synthesizer.

The DNA sequence to be used in producing the knockout construct is digested with a restriction enzyme selected to cut at a location(s) such that a new DNA sequence encoding a selectable marker gene may be inserted in the proper position within this DNA sequence. The proper position for a selectable marker gene insertion is that which will serve to modulate expression of the native gene. This position will depend on various factors such as the restriction sites in the sequence to be cut, and whether, for example, an intron sequence, an exon sequence or a promoter sequence is (are) to be modulated. In other words, the precise location of insertion of the selectable marker into the DNA sequence is that which will result in the modulation of promoter function or of synthesis of the native exon. For example, the knockout construct may be engineered to insert the selectable marker entirely within a single intron of the target gene. In this manner, the first nucleic acid sequence would comprise a region of the selected intron upstream from the second nucleic acid sequence and the second nucleic acid sequence would be selected comprising a region of the selected intron located downstream of the first nucleic acid sequence. The selectable marker would be introduced between the first and second nucleic acid sequences. When the construct is then introduced to the cell, the construct nucleic acid sequence is integrated into the target gene and the selectable marker is inserted entirely within the targeted intron.

Similarly, the construct may be engineered to insert the selectable marker within any desired and suitable region of the gene provided that expression is modulated. For example, the construct may be engineered to insert the selectable marker between two adjacent introns and thereby completely remove an endogenous exon of the target gene, to span over a region comprising at least a portion of an intron and at least a portion of an adjacent intron of the targeted gene, to span over a region comprising at least a portion the promoter for the targeted

gene to an adjacent intron, to span over a region encompassing more than one targeted gene, and combinations thereof.

After the genomic DNA sequence has been digested with the appropriate restriction enzymes, the selectable marker gene is ligated into the genomic DNA sequence using methods 5 well known to the skilled artisan and described in Sambrook *et al.*, *supra*. The ends of the DNA fragments to be ligated must be compatible; this is achieved by either cutting all fragments with enzymes that generate compatible ends, or by blunting the ends prior to ligation. Blunting is done using methods well known in the art, such as for example by the use of Klenow fragment (DNA polymerase I) to fill in sticky ends.

10 The ligated knockout construct may be introduced directly into the target cell, or it may first be placed into a suitable vector for amplification prior to insertion. Preferred vectors are those that are rapidly amplified in bacterial cells such as the pBluescript II SK vector (Stratagene, San Diego, Calif.) or pGEM7 (Promega Corp., Madison, Wis.).

In another embodiment of the invention, embryonic stem (ES) cells are used as the target 15 cell for their ability to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. For example, one mouse strain that is has been used for production of ES cells is the 129J strain. The cells are cultured and prepared for DNA insertion using methods well known to the skilled artisan such as those set forth by 20 Robertson (Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E. J. Robertson, ed. IRL Press, Washington, DC (1987)), Bradley *et al.* (*Current Topics in Devel. Biol.*, 20:357-371 (1986)), and Hogan *et al.* (Manipulating the Mouse Embryo: A Laboratory Manual, Cold

Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1986)), all of which are incorporated by reference herein.

Insertion of the knockout construct into the target cells may be accomplished using a variety of transfection methods well-known in the art. For example, suitable transfection 5 methods include electroporation, microinjection, and calcium phosphate treatment (see Lovell-Badge, in Robertson, ed., *supra*). A preferred method of transfection is electroporation. If the cells are to be electroporated, the targeted cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the cells are allowed to recover under suitable incubation conditions. 10 The cells are then screened for the presence of the knockout construct.

Each knockout construct DNA to be introduced into the cell must first be linearized if the knockout construct has been inserted into a vector. Linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

15 For introduction of the DNA sequence, the knockout construct DNA is added to the target cells under appropriate conditions for the insertion method chosen. Where more than one construct is to be introduced into the target cell, DNA encoding each construct may be introduced simultaneously or one at a time.

Screening may be done using methods known in the art or combinations thereof. Where 20 the selectable marker gene is an antibiotic resistance gene, the cells are cultured in the presence of an otherwise lethal concentration of the antibiotic. Those cells that survive have presumably integrated the knockout construct. If the selectable marker gene is other than an antibiotic resistance gene, the genomic DNA of the target cell may be extracted from the cells using

standard methods such as those described by Sambrook *et al.*, *supra*. The DNA may then be probed on a Southern blot with a probe or probes designed to hybridize only to the selectable marker sequence. If the selectable marker gene is a gene that encodes an enzyme whose activity may be detected (e.g., beta-galactosidase), the enzyme substrate may be added to the cells under 5 suitable conditions, and an appropriate assay for enzymatic activity may be conducted. In addition, the genomic DNA may be amplified by polymerase chain reaction (PCR) with probes specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only those cells containing the knockout construct in the proper position will generate DNA fragments of the proper size). PCR may be used in detecting the presence of homologous recombination 10 (Kim and Smithies, (1988) Nucleic Acid Res. 16:8887-8903; Joyner et al (1989) Nature 338:153-156). Primers may be used which are complementary to a sequence within the construct and complementary to a sequence outside the construct and at the target locus. In this way, one may only obtain DNA duplexes having both of the primers present in the 15 complementary chains in which homologous recombination has occurred. By demonstrating the presence of the primer sequences or the expected size sequence, the occurrence of homologous recombination is supported.

Upstream and/or downstream from the target gene knockout construct may be inserted a gene which provides for identification of whether a double crossover has occurred. For this purpose, any suitable marker may be used for as described herein. Preferably, the selectable 20 marker used to identify double crossovers is different than the selectable marker used to identify the integration of the target gene knockout construct. In one preferred embodiment, the herpes simplex virus thymidine kinase gene is employed, since the presence of the thymidine kinase gene may be detected by the use of nucleoside analogs, such as Acyclovir or Gancyclovir, for

their cytotoxic effects on cells that contain a functional HSV-tk gene. The absence of sensitivity to these nucleoside analogs indicates the absence of the thymidine kinase gene and, therefore, where homologous recombination has occurred, a double crossover event has also occurred.

The knockout construct may be integrated into several locations in the target cell genome, 5 and may integrate into a different location in each cell's genome, due to the occurrence of random insertion events. Notwithstanding random multiple integration sites, the desired location of the insertion is in a complementary position to the DNA sequence to be knocked out. It has been found that less than about 1-5% of the targeted cells that take up the knockout construct will actually integrate the knockout construct in the desired location. Identification of those cells 10 with proper integration of the knockout construct is described herein.

In one embodiment of the present invention, suitably transfected target cells containing the knockout construct in its proper location are inserted into an embryo. Insertion may be accomplished in any suitable method known in the art. Preferably, the cells are introduced into the embryo by microinjection. Most preferably, the cells are ES cells for injection into mouse 15 embryos. For microinjection, about 10-30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to integrate the transfected cell into the developing embryo. The suitable stage of development for injecting into the embryo is prior to the formation of the germinal layer of the developing embryo as one having ordinary skill in the art may readily determine. Preferably, the embryo is in the early blastocyst stage. By way of 20 example, mice embryos may be introduced to the transfected cells in about 3.5 days. The embryos are obtained by perfusing the uterus of pregnant females by methods known to the skilled artisan (e.g., Bradley (in Robertson, ed., *supra*)). Preferably, the embryos are male.

After the transfected target cell having proper integration of the target gene has been introduced into the embryo, the embryo is implanted into the uterus of a pseudopregnant foster mother. While any foster mother may be used, selection of the foster mother is based upon its ability to breed and reproduce well, and to care for its young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and is species dependent. For mice, this stage is about 2-3 days pseudopregnant.

In another embodiment, the suitable transfected target cells are nuclear transfer donor cells. Nuclear transfer donor cells may be virtually any somatic cell type and include fibroblasts, epithelial cells, cumulus cells, *etc.* Nuclear transfer donor cells are cultured *in vitro* and targeted using the constructs and techniques described herein via homologous recombination. Cells are grown in the appropriate medium to allow for selection of cells comprising the having properly integrated the knockout construct. PCR may also be done for confirmation of correctly targeted integration. Thereafter, an unfertilized oocyte of an animal is enucleated using known methods. The enucleated unfertilized oocyte is then fused to the selected knockout nuclear transfer donor cell. Fusion may be conducted by electrical stimulation, chemical stimulation, insertion by injection, or other known methods. The fused product is then cultured, assessed for viability and transferred to a surrogate recipient female. For reference and methods, see *e.g.*, Campbell *et al.* (1996) *Nature* 380:64; Wilmut *et al.* (1997) *Nature* 385:810; WO00/25578; WO97/07669; WO99/36510; WO00/42174; WO99/53751; WO99/45100, which are incorporated herein by reference.

Offspring or progeny that are born to the foster mother or surrogate recipient female are screened (*e.g.*, by PCR) for genomic DNA comprising the knockout construct. This step is

particularly important for selecting for progeny of foster mothers that carried embryos in which the transfected target cell was injected. On the other hand, the progeny of surrogate recipient females that carried the transfected target cell fused with the enucleated unfertilized oocyte will typically have the knockout construct inserted into its genome.

5 Any suitable selection method may be used. For example, if a coat color selection strategy has been used, the offspring may be screened for a coat color indicative of proper integration of the targeted gene into the offspring. Other methods include obtaining DNA from the offspring and screening for the presence of the knockout construct using Southern blots and/or PCR as described herein. Other means of identifying and characterizing the knockout 10 offspring include the use of Northern blots and Western blots. For example, Northern blots may be used to probe the mRNA for the presence or absence of transcripts encoding either the gene knocked out, the marker gene, or both. In addition, Western blots may be used to assess the level of expression of the gene knocked out in various tissues of these offspring by probing the Western blot with an antibody against the protein encoded by the gene knocked out, or an 15 antibody against the marker gene product, where this gene is expressed. *In situ* analysis (such as fixing the cells and labeling with antibody) and/or fluorescence activated cell sorting (FACS) analysis of various cells from the offspring may be conducted using suitable antibodies to look for the presence or absence of the knockout construct gene product.

Offspring that appear to contain the integrated knockout construct in its genome may then 20 be out-crossed to generate multiple offspring if they are believed to carry the knockout construct in their germ line to generate F1 offspring heterozygous for the knockout construct. F1's will then be crossed to generate homozygous knockout animals.

The heterozygotes may then be crossed with each other to generate homozygous knockout offspring. Homozygotes may be identified by any screening method as described herein. For example, the homozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from the host animal(s) that is (are) the product of this cross, as well 5 as host animals that are known heterozygotes and wild-type host animals. Probes to screen the Southern blots may be designed as set forth herein.

The knockout mammals described herein will have a variety of uses depending on the gene or genes that have been modulated. Where the targeted gene or genes modulated encode proteins believed to be involved in immunosuppression or inflammation, the knockout mammal 10 may be used to screen for drugs useful for immunomodulation, *i.e.*, drugs that either enhance or inhibit these activities. Screening for useful drugs may involve administering the candidate drug over a range of dosages to the knockout mammal, and assaying at various time points for immunomodulatory effects of the drug on the immune disorder being evaluated. Such assays may include, for example, looking for increased or decreased T and B cell levels, increased or 15 decreased immunoglobulin production, increased or decreased levels of chemical messengers such as cytokines (e.g., interleukins and the like), and/or increased or decreased levels of expression of particular genes involved in the immune response.

For example, patients undergoing chemotherapy often experience immunosuppression. It would be desirable to activate the immune system of such individuals by administering to the 20 patient a therapeutic agent capable of producing such an effect. A knockout mammal as described herein could be used to screen a variety of compounds, either alone or in combination, to determine whether partial or total restoration or activation of the immune response results.

Similarly, the same strategy could be applied to find compounds that would be useful in suppressing the inflammatory response observed in many patients with arthritis, or useful in suppressing the autoimmune phenomenon observed in patients with rheumatoid arthritis and lupus. In addition, mammals may be useful for evaluating the development of the immune system, and for studying the effects of particular gene mutations.

5 In a preferred embodiment, the knockout mammals described herein are used for xenograft transplantation into human patients. The xenograft tissue may be from any mammal, preferably a pig. The xenotransplanted tissue may be in the form of an organ including, for example, a kidney, a heart, a lung, or a liver. Xenotransplant tissue may also be in the form of 10 parts of organs, cell clusters, and glands including, for example, lenses, pancreatic islet cells, skin, corneal tissue, and the like.

In yet another aspect of the present invention, the target gene is the Gal α (1,3) galactosyl transferase gene in pigs. The Gal α (1,3) galactosyl transferase is an attractive target for knockout in the pig. This enzyme is responsible for the addition of a carbohydrate residue, 15 Gal α (1,3) Gal, that is recognized by human IgM and IgG antibodies in pig-to-human xenotransplanted tissue and leads to subsequent hyperacute rejection. Knockout pigs, which lack the Gal α (1,3) galactosyl transferase gene, may thus potentially serve as a rich source for xenotransplanted organs. Nucleic acid sequences encoding Gal α (1,3) galactosyl transferase and mutants thereof are disclosed. Preferably, the nucleotide sequence encodes pig Gal α (1,3) 20 galactosyl transferase. Nucleotide sequences may be in the form of DNA, RNA or mixtures thereof. Nucleotide sequences or isolated nucleic acids may be inserted into replicating DNA, RNA or DNA/RNA vectors (as are well known in the art), such as plasmids, viral vectors, and

the like (Sambrook et al, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY, Second Edition 1989).

Nucleotide sequences encoding Gal α (1,3) galactosyl transferase may include promoters, enhancers and other regulatory sequences for expression, transcription and translation. Vectors 5 encoding such sequences may include restriction enzyme sites for the insertion of additional genes and/or selection markers, as well as elements necessary for propagation and maintenance of vectors within cells.

Targeting constructs comprising nucleotide sequences, and mutants thereof, of the Gal α (1,3)galactosyl transferase are particularly preferred as they may be used to inactivate wild type 10 Gal α (1,3) galactosyl transferase genes according to the methods of the present invention. Mutant Gal α (1,3) galactosyl transferase nucleotide sequences include, but are not limited to, nucleotide deletions, insertions, substitutions and additions to wild type Gal α (1,3) galactosyl transferase, such that the resultant mutant does not encode a functional galactosyl transferase. These nucleotide sequences may be utilized in the methods of modulating expression of 15 galactosyl transferase of the present invention. In this manner, mutant sequences are recombined with wild type genomic sequences in the target cells.

In a most preferred embodiment, knockout pigs are produced in which the Gal α (1,3)galactosyl trasferase gene produces a non-functional protein. By producing a non-functional protein, the human antibody that would otherwise bind to the Gal α (1,3)Gal epitope 20 expressed on the xenotransplanted tissue does not bind, so that immune responses which give rise to tissue rejection are prevented. In this embodiment, any knockout construct capable of

modulating the interaction between antibodies directed to the Gal α (1,3) galactosyl transferase linkage may be used.

EXAMPLES

The following examples are for purposes of illustration only, and are not intended to limit
5 the scope of the disclosure or claims.

EXAMPLE 1

Inactivation of the Gal α (1,3) galactosyl transferase gene by insertion of an engineered active gene in the form of an engineered exon within intron 3.

In this example, the Gal α (1,3) galactosyl transferase protein is prevented from being
10 translated by insertion of an in-frame, promoterless engineered exon (e.g., an antibiotic resistance gene) that contains multiple stop codons within an intron of the Gal α (1,3) galactosyl transferase gene. Using this 'promoter-trap' strategy, the engineered exon is spliced in frame upstream of exon 4 of the Gal α (1,3) galactosyl transferase gene. This results in the expression of the drug resistance gene prior to the gene of interest and concomitantly inhibits expression of the transferase gene due to the
15 presence of multiple stop codons downstream of the drug resistance gene. As described herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon, including, but not limited to, ampicillin, kanamycin, gentamicin, neomycin phosphotransferase, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. The present example employs neomycin.

20 A gene targeting construct is designed which contains a sequence with homology to the Gal α (1,3) galactosyl transferase gene 5' intron 3 sequence, an intron 4 splice acceptor signal sequence, a promoterless neomycin phosphotransferase gene engineered to contain multiple stop

codons (engineered exon), the intron 4 splice donor sequence for splicing the engineered exon to the downstream exon 4, and additional intron 3 sequence homology to aid with annealing to the porcine Gal α (1,3) galactosyl transferase gene. Although this example describes targeting intron 3, it will be appreciated that the method may be used to target any intron within the Gal α (1,3) galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal α (1,3) galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1. A schematic diagram of the targeting vector and corresponding nucleotide sequence are shown in Figures 2 and 3.

The gene targeting construct is generated by ligating 5 distinct DNA fragments (1-5 below) together to form the final gene targeting construct using standard molecular biology techniques well known to those skilled in the art. The PCR reactions use the ELONGASE Enzyme Mix (Life Technologies, Gaithersburg, MD) according to the manufacturer's instructions. In the present example, a 50 μ l final reaction volume is used, with 2 μ l of DNA template, 1 μ l of ELONGASE Enzyme Mix, 60 mM Tris-SO₄ (pH9.1) 18mM (NH₄)₂SO₄, 1.2 mM MgSO₄, 200mM dNTP mix, 10% DMSO and 200nM of each primer. The reaction is hot started at 95°C for 1 minute and followed by 30-40 cycles in a standard PCR thermocycler (GeneAmp PCR System 2400; PE Applied Biosystems, Foster City, CA).

1. A polymerase chain reaction (PCR) product consisting of intron 3 sequences as listed in Figures 1 and 3, nucleotide numbers 10-4020, is generated using standard PCR conditions for 20 long range PCR of genomic fragments. Primers used include a 5' primer containing a NotI restriction site and intron 3 sequences 10-23 (GGCGGCCGCAGGCCTCACTGGCC); and a 3' primer containing a SalI restriction site and sequences homologous to intron 3 sequences 3999-4020 (GGTCGACGGATGCTGGGTGGAATAACAGG), where underlined sequence

indicates restriction sites and bold type indicates homology to endogenous sequences. An additional guanine nucleotide is added to the 5' end of all probes in this example to balance out 1 bp deletions that sometimes occur during cloning.

2. A PCR product is generated consisting of intron 4, the 3' splice sequence (the pyrimidine 5 rich lariat and Gal α (1,3) galactosyl transferase intron 3 dinucleotide acceptor sequences) and 196 bases 5' flanking the **ag** dinucleotide acceptor site (nucleotides 11521-11716 in Figures 1 and 3). Primers used include a 5' primer containing a Sall restriction site (**GGTCGACCCACCGTTGATCTGAG**); and a 3' primer containing a EcoRI restriction site and the complementary strand homologous to the pyrimidine rich lariat and Gal α (1,3) galactosyl transferase dinucleotide acceptor sequences (GGAATTCCTAAAGCAAATGGAAATAAACATATC), where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

3. A PCR product consisting of a neomycin resistance gene (Genbank Accession #AF081957; Figure 4) is generated using a 5' primer containing an EcoRI restriction site, and homology to the neomycin resistance gene, including the ATG start codon (**GGAATTCAATGGATCCCCACCATGG**); and a 3' primer containing a HindIII restriction site and complementary strand sequences to the 3' coding region of the neomycin gene, including the natural stop codon followed by two additional engineered stop codons (GAAGCTTCGGCTATTACTAAGTAGTGGATATCC), where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence (see Figures 3 and 4).

4. A PCR product is generated containing the 5' splice donor sequences for intron 4 of the Gal α (1,3) galactosyl transferase gene, corresponding to sequences 4938-5173 in the claimed sequence comprising intron 4 (Figures 1 and 3). Primers used include a 5' primer containing a HindIII site and sequence identity to intron 4, sequences 4938-4962, including the Gal α (1,3) galactosyl transferase dinucleotide splice site (**GAAGCTTGTAAATTATGAAACATGATG**); and a 3' primer containing a PstI site and complementary strand sequence from intron 4 corresponding to nucleotide numbers 5152-5173 and includes multiple stop codons (**GCTGCAGCCACAGGTCACGGCAATGCGG**); where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

5 10 5. A PCR product containing 1150 nucleotides of intron 3, corresponding to nucleotides 4024-4826 of the claimed sequence (Figures 1 and 3). Primers used include a 5' primer containing a PstI site and sequences 4024-4050 of the claimed sequence (**GCTGCAGCCCTCTTCAACTACAATTTCATGCAGC**); and a 3' primer containing a XhoI restriction site and complementary strand sequences to 4801-4826 of the claimed sequence 15 (**GCTCGAGAGAAAATTAGATTAAATACACCCAGAG**); where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

Each PCR fragment (steps 1-5) is separately amplified. A single PCR fragment is cloned into the pCR2.1 vector (Invitrogen, San Diego, CA) according to the manufacturer's ligation instructions. The recombinant plasmid DNA is transformed into a suitable bacterial host (Invitrogen, San Diego, CA). The bacteria are cultured and plasmid DNA is isolated. Plasmid DNA with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.

Following PCR fragment amplification, a series of ligations is performed to clone the final construct in the bacterial plasmid pBS SK+ (Stratagene, La Jolla, CA).

- a. The HindIII-PstI fragment from the intron 4 PCR product (step 4) and the PstI-XhoI 3' homology fragment from intron 3 (step 5 above) are ligated to a pBS SK+ vector DNA following 5 digestion with HindIII and XhoI. The 3 DNA fragments are mixed in equal molar ratios and incubated in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life Technologies, Gaithersburg, MD). The bacteria are cultured, and plasmid DNA is isolated. Plasmid with the correct insert, as 10 determined by restriction analysis and sequence analysis, is used to construct the final product.
- b. The resulting plasmid (step 5a) is digested with HindIII and EcoRI and ligated with the HindIII-EcoRI Neomycin resistance gene fragment (step 3), that has been previously digested with HindIII and EcoRI. The resulting recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life Technologies, Gaithersburg, MD). The bacteria are cultured, and 15 plasmid DNA is isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.
- c. The resulting plasmid (step 5b) is digested with EcoRI and NotI and ligated to the SalI-EcoRI intron 4-3' splice fragment (step 2) previously digested with SalI and EcoRI and the intron 3 4 kb NotI-SalI fragment (step 1) previously digested with NotI and SalI. The 3 DNA 20 fragments are incubated in equal molar ratios in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life

Technologies, Gaithersburg, MD). The bacteria are cultured, and the recombinant plasmid DNA is isolated.

This final construct is used to transfect porcine embryonic fibroblasts, transgenic pig fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones that 5 are resistant to neomycin are screened by PCR to determine the site of integration. A primer located in the region of intron 4 not incorporated into the final construct (complementary strand of 5407-5427; GGACAATGGCAACATGGCAGG; see Figures 1 and 3) is used in combination with the 5' neomycin gene primer (step 3). Only targeted insertions yield the appropriate sized PCR fragment. All other integration events produce a negative result.

10 Cell clones with a targeted insertion are then used to generate transgenic animals using nuclear transfer techniques, or in the case of the stem cells, used to inject into developing blastocysts and produce chimeric offspring.

EXAMPLE 2

Inactivation of the Gal α (1,3) galactosyl transferase gene by replacement of exon 4 with an active 15 gene in the form of an engineered exon.

In this example, the Gal α (1,3) galactosyl transferase protein was prevented from being translated by replacing an endogenous exon (exon 3) with an in-frame, promoterless engineered exon (an antibiotic resistance gene) that contained a bovine growth hormone poly A sequence attached to the 3' end of the gene, which served to terminate transcription of the engineered exon. 20 The engineered exon was spliced in frame, so as to take advantage of the endogenous promoter typically used by the Gal α (1,3) galactosyl transferase gene ('promoter-trap' strategy). This resulted in the expression of the drug resistance gene and concomitantly inhibited expression of the full-length

Gal α (1,3) galactosyl transferase gene. As described herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon, including, but not limited to, ampicillin, kanamycin, gentamicin, neomycin phosphotransferase, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase.

5 The present example utilizes puromycin.

A gene targeting construct was designed which contained a sequence with homology to the Gal α (1,3) galactosyl transferase gene 3' intron 3 sequence, an intron 3 splice signal sequence (splice acceptor sequence), a Kozak consensus sequence, a promoterless puromycin N-acetyl transferase gene linked to a bovine growth hormone poly A sequence (bpoly A) (engineered exon),
10 the 5' intron 4 splice signal sequence (splice donor sequence), and a sequence with 5' intron 4 sequence homology. Exon 4 of the Gal α (1,3) galactosyl transferase gene codes for ATG start codon and the N-terminal portion of the protein. Although this example describes targeting introns 3 and 4, it will be appreciated that the method may be used to target any exon within the Gal α (1,3) galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal α (1,3)
15 galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1.

The gene targeting construct was generated by ligating two distinct DNA fragments together to form the final gene targeting construct using standard molecular biology techniques well known to those skilled in the art. The first DNA fragment was obtained from the 3' end of intron 3 containing the 3' splice sequence (the pyrimidine-rich branch site used in forming the
20 lariat during splicing and the AG dinucleotide splice acceptor sequence). The second DNA fragment was obtained from the 5' end of intron 4 containing the GT dinucleotide splice donor sequence. The fragments were ligated into the pBluescript vector containing a Kozak consensus sequence in-frame with the coding sequence of a promoterless puromycin gene linked to the

bovine growth hormone poly A sequence (Figure 5) to form the final gene construct. A schematic diagram of the targeting vector and corresponding nucleotide sequence are shown in Figures 6 and 7. Additionally, this construct has been deposited with ATCC on 28 September 2000 with accession number _____.

5 1. Generation of the first DNA fragment.

The first DNA fragment was a polymerase chain reaction (PCR) product consisting of intron 3 sequence as shown in Figure 7 (nucleotide numbers 235-4851, positions relative to nucleotide position 1 of the insert isolated from the lambda phage clone) and generated using standard PCR conditions as described by Randolph *et al.*, (1996) for long range PCR of genomic 10 fragments. The 5' primer, consisting of intron 3 sequences 235-260, was 5'-AAGATTATAAATAGCCTCGTGTCAAGG-3'. The 3' reverse primer sequence was complementary to sequence 4827-4851 at the extreme 3' end of intron 3 and containing the AG splice acceptor site, and was 5'-CTCCTGGGAAAAGAAAAGGAGAAGG-3'.

PCR reaction conditions to generate the 4.616 kb intron 3 sequence were performed using 15 the ELONGASE Enzyme Mix (Life Technologies, Gaithersburg, MD) according to manufacturer's conditions. In the present example, a 50 μ l final reaction volume was used, with 2 μ l of DNA template, 1 μ l of the ELONGASE Enzyme Mix, 60 mM Tris SO₄ (pH 9.1), 18 mM (NH₄)₂SO₄, 1.2 mM MgSO₄, 200 mM dNTP mix, 10% DMSO and 200 nM of each primer. The reaction was hot started at 95°C for 1 min, followed by 30-40 cycles in a standard PCR machine 20 (e.g., Gene Amp PCR Systems 2400; PE Applied Biosystems, Foster City, CA).

2. Preparation of the PCR2.1 cloning vector.

The Not I site of the PCR2.1 cloning vector (Invitrogen, San Diego, CA) was destroyed to avoid carrying over a second Not I site into the final construct. The Not I site was unique and

used to linearize the final plasmid construct. The PCR2.1 vector was digested with Not I and the overhangs filled-in using the Klenow enzyme (Roche Molecular Biochemicals, Indianapolis, IN) according to the manufacturer's specifications. The plasmid was re-ligated using T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations.

5 Plasmid DNA was transformed into a suitable bacterial host (Top 10 F', Invitrogen, San Diego, CA). The bacteria were cultured, and plasmid DNA was isolated and incubated with Not I enzyme to confirm loss of this site by restriction analysis.

3. Insertion of the first DNA fragment into the PCR2.1 vector.

Following PCR, the 4.616 kb fragment was ligated into the modified PCR2.1 vector 10 using T4 DNA ligase according to the manufacturer's specifications. Plasmid DNA was transformed into a suitable bacterial host (e.g., Top 10 F', Invitrogen, San Diego, CA). The bacteria are cultured and plasmid DNA is isolated. Plasmid with the correct insert in the proper orientation, as determined by restriction analysis and sequence analysis, was used to construct the final product.

15 4. Preparation of the second DNA fragment.

A 2.084 kbp PCR product consisting of the intron 4 homology sequence containing the GT dinucleotide donor consensus splice sequence was constructed using standard PCR conditions as described above in step 1. The 5' primer consisting of sequence 4938-4961 at the extreme 5' end of intron 4 was 5'-GTAATTATGAAACATGATGAAATG-3'. The 3' primer 20 was homologous to the complementary strand of intron 4 at position 6997-7021 and has the sequence 5'-AGCCAGCGCTTACTAAGTACGTTGC-3'

5. Insertion of the second DNA fragment into the PCR2.1 vector.

Following PCR, the 2.084 kb fragment was ligated into the pCR2.1 vector from Invitrogen (San Diego, CA) using the manufacturer's ligation conditions. Following ligation, the recombinant plasmid DNA was transformed into a suitable bacterial host (Top 10 F', Invitrogen, San Diego, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the 5 correct insert, and orientation in the plasmid, as determined by restriction analysis and sequence analysis, was used to construct the final product.

6. Preparation of a synthetic oligonucleotide linker sequence.

A synthetic oligonucleotide linker containing a Kozak consensus sequence and relevant restriction enzyme sites was prepared for in-frame cloning of the promoterless puromycin gene:

10 Xho I | Kozak seq.--- | HpaI | Hind III | Bgl II | Sal | Eco RV | Eco RI
TCGAGCCACCATGGTTAACAAAGCTTAGATCTGTCGACGATATCG
CGGTGGTACCAATTGTTGAATCTAGACAGCTGCTATAGCTTAA

7. Assembly of the gene targeting construct.

The following ligations were performed to generate the final construct in the bacterial 15 plasmid pBS KS+ (Strategene, La Jolla, CA). The final construct is illustrated in Figure 6:

a. The oligonucleotide linker containing the Kozak consensus sequence (step 6) was ligated to the pBS KS+ vector DNA following digestion with Xho I and Eco RI. Ligation was carried out using at least a 3:1 molar ratio of linker to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following 20 ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Restriction enzyme analysis was performed to confirm successful ligation using unique

restriction sites within the linker (Bgl II or Hpa I). This plasmid containing the linker was then used to construct the final product.

b. The resulting "mother" plasmid (step 7a) was then digested with Eco RV and Spe I to clone in the 3' arm of the targeting construct. The 2.084 kb PCR fragment cloned into the PCR 5 2.1 vector (step 5) was digested with Eco RV and Spe I, isolated away from vector DNA by agarose gel electrophoresis and purified. The 2.084 kb fragment was ligated between the EcoRV and Spe I sites of the mother plasmid (step 7a). Ligation was carried out using a 3:1 molar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid 10 was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene , La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was then used to construct the final product.

c. The next fragment cloned into the mother plasmid (step 7b) was the cassette containing 15 the promoterless puromycin gene coding sequence with the bovine growth hormone gene polyA signal sequence attached to its 3' end following the TGA stop codon (Figure 5). The PGK puromycin bpolyA plasmid (used as a positive control for puromycin resistance of transformed cells) was digested with Hind III and Xho I. The puromycin bpolyA fragment was separated away from the rest of the vector DNA containing the PGK promoter by electrophoresis on a 20 0.7% agarose gel and purified. The mother plasmid (step 7b) was digested with Hind III and Sal I. The Hind III/Xho I puromycin bpolyA cassette was ligated to the Hind III and Sal I sites of the mother plasmid. Ligation was carried out using a 3:1 molar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the

manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was used to construct the final product.

5 d. The final cloning step involved ligating the 5' arm of the construct, which was the 4.616 kb intron 3 insert from the PCR2.1 vector (step 3). The PCR2.1 vector (step 3) was digested with Kpn I and Xho I. The 4.616 kb PCR fragment was isolated away from vector DNA by agarose gel electrophoresis and purified. The 4.616 kb Kpn I/Xho I insert was ligated into the mother plasmid (step 7c) that was digested with Kpn I and Xho I. Ligation was carried out using
10 equimolar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated by standard molecular biology techniques. Plasmid with the correct insert, as determined by
15 restriction analysis and sequence analysis, was used as the final product.

e. The final construct may be used to transfect porcine embryonic fibroblasts, transgenic porcine fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones that are resistant to puromycin may be screened by PCR to determine the site of integration by methods well known to those of skill in the art. A primer located in a region of
20 intron 4, which is not incorporated into the final construct, may be used in combination with a 5' puromycin gene primer. Only targeted insertions will yield the appropriate size PCR fragment. All other integration events will produce a negative result.

Cell clones with a targeted insertion may then be used to generate transgenic animals using nuclear transfer techniques, or in the case of stem cells, used to inject into developing blastocysts and produce chimeric offspring.

Example 3

5 Inactivation of the Gal α (1,3) galactosyl transferase gene by replacement of exon 4 with an a reverse orientation active gene.

In this example, the Gal α (1,3) galactosyl transferase gene was functionally inactivated by using a "collision construct" to insert an active gene in place of an exon and at least part of the flanking introns, including the splice donor and splice acceptor sites. The inserted gene is under the 10 control of a highly active promoter such as the phosphoglycerate kinase I (PGK) gene promoter, such that transcription of this gene causes the termination of transcription of the endogenous gene (Rosario *et al.*, (1996) *Nat. Biotech.* 14:1592-1596). Exon 4 of the Gal α (1,3) galactosyl transferase gene codes for ATG start codon and the N-terminal portion of the protein. Thus, the insertion was made to replace exon 4 as well as a portion of the flanking introns 3 and 4, resulting in a truncated transcript 15 that did not code for a functional enzyme. Although this example describes targeting introns 3 and 4, this method could be used to target any introns within the Gal α (1,3) galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal α (1,3) galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1.

In this example, the PGK promoter was inserted driving the expression of the puromycin 20 resistance gene with the bovine growth hormone poly A (bpolyA) transcription termination sequence. This gene replaced the Gal α (1,3) galactosyl transferase exon 4 as well as a portion of the flanking intron 3 and 4 sequences by standard homologous recombination techniques utilizing intron 3

and 4 sequences for homology flanking the inserted gene. Intron 3, which separates exons 3 and 4, is greater than 5 kb in length, and a construct was built such that there was at least about 4.6 kb of homologous sequence on one end of the gene. Intron 4, which separates exons 4 and 5 is about 6.8 kb in length, and the construct was built such there is at least about 2.2 kb of homologous sequence on 5 the other end of the gene. Positive selection for transfected cells in which the construct has been integrated was accomplished via expression of the puromycin resistance gene. As described herein, it will be appreciated that any selection marker gene that confers survival of the targeted cells under appropriate selection conditions may be driven by the strong PGK promoter. Additionally, a toxin gene was inserted to eliminate random integration events.

10 The collision construct was generated using standard molecular biology techniques well known to those skilled in the art. The 4.616 kb intron 3 homology fragment and the 2.084 kb intron 4 homology fragment were generated using PCR and cloned into the PCR2.1 cloning vector as described in Example 2, steps 1-5 above for the replacement targeting construct. The generation of the collision construct first involved ligating the 2.084 kb intron 4 homology 15 fragment into the pBS KS+ vector as the 3' arm of the collision construct, followed by the PGK-puromycin-bovine polyA cassette in the opposite orientation to the coding sequence of the GT gene. The 4.616 kb intron 3 homology fragment, as the 5' arm, was cloned in next. This generated the targeting construct for homologous recombination. The ricin A toxin gene was also added to the plasmid outside the region of homology, which will effectively kill a 20 percentage of the cells in which random integration has occurred. The ricin A toxin gene was PCR amplified and cloned based upon the published sequence (Figure 8). A schematic diagram of the final construct is shown in Figure 9. Additionally, this collision construct has been deposited with ATCC on 28 September 2000 with accession number _____.

1. The 2.084 kb 3' arm of the construct (intron 4 homology fragment) was the first fragment to be ligated into the pBluscript cloning vector, which was modified to contain the XhoI – EcoRI linker (see Example 2, step 6) within its multiple cloning site (pBS KS+). The ligation of the linker into the vector is described above in Example 2, step 7a, and ligation of the 2.084 kb 3' linker into the Eco RV and Spe I sites of the vector is described above in Example 2, step 7b.
5
2. The next step involved ligation of the PGK puromycin bovine polyA cassette into the pBS KS+ vector, which contained the 2.084 kb 3' arm. The PGK-puro-bPA cassette was digested with Eco RI which was immediately 5' of the PGK promoter. The Eco RI overhangs were blunted by filling in with Klenow enzyme (Roche Molecular Biochemicals, Indianapolis, IN) using the manufacturer's specifications. The PGK-puro-bPA cassette was then released from the vector by digestion with Xho I, which was immediately 3' of the bovine polyA sequence. The blunted PGK-puro-bPA-Xho I cassette was separated from vector DNA by agarose gel electrophoresis and purified. The pBS KS+ vector (step 1) was digested with Hpa I and Xho I, and the blunted PGK-puro-bPA-Xho I fragment was ligated between the Hpa I and Xho I sites of the vector. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Stratagene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was then used as the final product.
10
15
3. The 4.616 kb intron 3 homology fragment was ligated into the pBS KS+ mother plasmid (step 2) and represented the 5' arm of the collision construct. Isolation of this fragment from the PCR2.1 cloning vector and ligation into the Kpn I and Xho I sites of the mother plasmid is described above in Example 2, step 7d.
20

4. The ricin A toxin gene (Figure 8) was inserted into a commercially available mammalian expression vector, *e.g.*, pcDNA1/Amp (Invitrogen). The insert was then excised with the CMV promoter and the SV40 poly A site and cloned into the Not I site of the recombinant plasmid by blunt end ligation, following Klenow fill in reactions on both the insert and vector.

5 5. Following ligation, the recombinant plasmid DNA was used to transform a suitable bacterial host (XLI-blue, Strategene). The bacteria were cultured, and plasmid DNA isolated. This final construct DNA was then linearized with Kpn I (a unique enzyme site in the plasmid MCS outside of the construct sequence). Linearized plasmid may be used to transfect porcine embryonic fibroblasts, transgenic porcine fibroblasts, or porcine embryonic stem cells, or porcine 10 primordial germ cells. Cell clones resistant to puromycin may be screened by PCR to determine the site of integration by methods well known to those of skill in the art. A primer located in the region of intron 4 not incorporated into the final construct may be used in combination with a 5' puromycin gene primer. Only targeted insertions yield the appropriate size PCR fragment. All other integration events produce a negative result.

15 Cell clones with a targeted insertion may then be used to generate transgenic animals using nuclear transfer techniques, or in the case of stem cells, used to inject into developing blastocysts and produce chimeric offspring.

EXAMPLE 4

Isolation of porcine genomic DNA encoding the Gal α (1,3) galactosyl transferase gene from a 20 Lambda phage clone library.

In this example, a pig genomic library was screened using a cDNA corresponding to exon 4 of the Gal α (1,3) galactosyl transferase gene using molecular biology techniques that are well

known to those skilled in the art (e.g., Sambrook *et al.*, *supra*). A pig genomic library (Clontech, Palo Alto, CA) was obtained and screened with a PCR fragment derived from exon 4 of the porcine Gal α (1,3) galactosyl transferase gene. Exon 4 was labeled with 32 P dCTP using the Random Prime Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. Approximately 4 5 million phage forming units were screened and unique clones that contain exon 4 sequences as determined by Southern blotting were isolated. Clones obtained by this procedure contained inserts 15-40kb in length. These clones, designated pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2. Five vectors comprising unique, overlapping nucleotide sequences which span the entire the pig Gal α (1,3) galactosyl transferase gene from within intron 3 through intron 8 have 10 been deposited with the ATCC: (1) a 1.6 kb insert within intron 3 of the extreme 5' end of the 18.275 kb lambda-2 phage clone, (2) a 6.7 kb HindIII fragment spanning intron 3 to intron 4 of the 18.275 kb lambda-2 phage clone, (3) a 4 kb HindIII fragment following the 6.7 kb fragment 2 of the 18.275 lambda-2 phage clone, (4) a 6 kb HindIII-SalI fragment at the 3' most portion of the 18.275 lambda-2 phage clone, and (5) a 13 kb fragment of the lambda-2 phage clone 15 spanning exon 7 to exon 9. These five vectors were deposited with ATCC on 29 September 2000 with accession numbers _____, respectively. Subclones of the various inserts were used to generate the claimed intron sequences from within intron 3 to intron 8 as provided in Figure 1 using molecular biology techniques well-known to those skilled in the art (see e.g., Sambrook *et al.*, *supra*). These sequences may be used to determine regions of 20 sequence homology in design of targeting constructs for modulation of the pig Gal α (1,3) galactosyl transferase gene

Although the compositions and methods provided herein have been set forth in detail, one skilled in the art will recognize that numerous changes and modifications may be made, and that

such changes and modifications may be made without departing from the spirit and scope thereof.

We Claim:

1. A method of modulating the expression of a eukaryotic gene in a cell, comprising
transfected said cell with a nucleic acid construct, said construct comprising a first
construct sequence homologous to a first gene sequence, a sequence encoding a selectable
marker, and a second construct sequence homologous to a second gene sequence, wherein said
first and second gene sequences independently comprise at least a portion of one or more intron
regions of said eukaryotic gene, and
integrating said selectable marker into said eukaryotic gene,
wherein expression of said selectable marker results in modulation of expression of said
eukaryotic gene in said cell.
- 10 2. The method of claim 1, wherein said first construct sequence and said second construct
sequence are each homologous to at least a portion of an intron region of the gene.
- 15 3. The method of claim 1, wherein said sequence encoding a selectable marker is integrated
into said eukaryotic gene by homologous recombination, wherein said first construct sequence
recombines with said first gene sequence and said second construct sequence recombines with
said second gene sequence to insert the selectable marker into the gene.
4. The method of claim 1, further comprising screening said cell for expression of said
selectable marker.
- 20 5. The method of claim 2, wherein said first construct sequence and said second construct
sequence are homologous to different regions from within the same intron of the gene.

6. The method of claim 2, wherein said first construct sequence and said second construct sequence are homologous to regions of different introns.
7. The method of claim 1, wherein said selectable marker gene is an antibiotic resistance gene.
- 5 8. The method of claim 1, wherein said sequence encoding a selectable marker is a nucleotide sequence which, when expressed, confers a phenotype selected from the group consisting of ampicillin resistance, kanamycin resistance, gentecin resistance, neomycin resistance, puromycin resistance, hygromycin b resistance, thymidine kinase activity, tryptophan synthetase activity, adenine phosphribosyltransferase activity, dihydrofolate reductase activity, 10 and histidinol dehydrogenase, anthocyanin, bets-glucuronidase and luciferase.
9. The method of claim 8, wherein said sequence encoding a selectable marker confers neomycin resistance or puromycin resistance.
10. The method of claim 1, wherein said eukaryotic gene is selected from the group consisting of genes encoding B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2, VCAM-1, CD28, 15 CD80, CD86, CD154, major histocompatibility complex class I β -2-microglobulin, invariant chain, caspase-1 caspase-3, and Gal α (1,3) galactosyl transferase.
11. The method of claim 10, wherein said eukaryotic gene encodes Gal α (1,3) galactosyl transferase.
12. The method of claim 11, wherein said Gal α (1,3) galactosyl transferase gene is a porcine 20 gene.

13. The method of claim 12, wherein said first construct sequence and said second construct sequence are independently selected from homologous regions of the intron selected from the group consisting of intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, and intron 9 of the porcine Gal α (1,3) galactosyl transferase gene.
- 5 14. The method of claim 13, wherein intron 4 has the nucleotide sequence of nucleotides 4938-11716 in Figure 1.
15. The method of claim 13, wherein intron 5 has the nucleotide sequence of nucleotides 11753-13748 in Figure 1.
- 10 16. The method of claim 13, wherein intron 6 has the nucleotide sequence of nucleotides 13810-14358 in Figure 1.
17. The method of claim 13, wherein intron 7 has the nucleotide sequence of nucleotides 14463-21627 in Figure 1.
18. The method of claim 13, wherein intron 8 has the nucleotide sequence of nucleotides 21766-27048 in Figure 1.
- 15 19. The method of claim 13, wherein said first construct sequence and said second construct sequence are homologous to different regions within the same intron of the eukaryotic gene.
20. The method of claim 19, wherein said intron is intron 3 of the porcine Gal α (1,3) galactosyl transferase gene.
21. The method of claim 13, wherein said first construct sequence and said second construct sequence are homologous to different introns of porcine Gal α (1,3) galactosyl transferase gene.

22. The method of claim 21, wherein said first construct sequence is upstream of said second construct sequence.
23. The method of claim 21, wherein said first construct intron region is homologous to an intron 3 region and said second construct intron region is homologous to an intron 4 region of 5 porcine Gal α (1,3) galactosyl transferase.
24. The method of claim 2, wherein said sequence encoding a selectable marker is a promoterless gene.
25. The method of claim 2, wherein said sequence encoding a selectable marker further comprises a promoter.

10 26. The method of claim 25, wherein said promoter is a phosphoglycerate kinase (PGK) promoter.

27. The method of claim 2, wherein said sequence encoding a selectable marker is transcribed in the opposite orientation relative to the orientation of said eukaryotic gene.

15 28. The method of claim 27, wherein said sequence encoding a selectable marker further comprises a promoter sequence.

29. The method of claim 1, wherein said cell is selected from the group consisting of a fibroblast, epithelial cell, endothelial cell, transgenic embryonic fibroblast, embryonic stem cell, and primordial germ cell.

30. The method of claim 2, wherein said cell is a porcine cell.
31. The method of claim 2, wherein said construct further comprises an AG dinucleotide splice acceptor site.
32. The method of claim 2, wherein said construct further comprises a GT dinucleotide splice 5 donor site.
33. A nucleic acid construct comprising a first construct sequence homologous to a first gene sequence, a sequence encoding a selectable marker, and a second construct sequence homologous to a second gene sequence, wherein said first and second gene sequences independently comprise at least a portion of one or more intron regions of a eukaryotic gene.
- 10 34. The nucleic acid construct of claim 33, further comprising an AG dinucleotide splice acceptor site.
35. The nucleic acid construct of claim 33, further comprising a GT dinucleotide splice donor site.
36. The nucleic acid construct of claim 33, further comprising a Kozak consensus sequence.
- 15 37. The nucleic acid construct of claim 33, wherein said sequence encoding a selectable marker is a nucleotide sequence, which when expressed, confers a phenotype selected from the group consisting of ampicillin resistance, kanamycin resistance, gentecin resistance, neomycin resistance, puromycin resistance, hygromycin b resistance, thymidine kinase activity, tryptophan synthetase activity, adenine phosphribosyltransferase activity, dihydrofolate reductase activity, 20 and histidinol dehydrogenase, anthocyanin, beta-glucuronidase and luciferase.

38. The nucleic acid construct of claim 37, wherein said sequence encoding a selectable marker confers puromycin resistance or neomycin resistance.

39. The nucleic acid construct of claim 33, wherein said eukaryotic gene is selected from the genes encoding B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2, VCAM-1, CD28, CD80, CD86, 5 CD154, major histocompatibility complex class I β -2-microglobulin, invariant chain, caspase-1, caspase-3, and Gal α (1,3) galactosyl transferase.

40. The nucleic acid construct of claim 39, wherein said eukaryotic gene is porcine Gal α (1,3) galactosyl transferase.

41. The nucleic acid construct of claim 40, wherein said first construct sequence and said 10 second construct sequence are independently selected from homologous regions of the intron selected from the group consisting of intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, and intron 9 of the porcine Gal α (1,3) galactosyl transferase gene.

42. The nucleic acid construct of claim 41, wherein intron 4 has the nucleotide sequence of nucleotides 4938-11716 in Figure 1, intron 5 has the nucleotide sequence of nucleotides 11753- 15 13748 in Figure 1, intron 6 has the nucleotide sequence of nucleotides 13810-14358 in Figure 1, intron 7 has the nucleotide sequence of nucleotides 14463-21627 in Figure 1, and intron 8 has the nucleotide sequence of nucleotides 21766-27048 in Figure 1.

43. The nucleic acid construct of claim 41, wherein said first construct sequence and said 20 second construct sequence are homologous to different regions within the same intron of the eukaryotic gene.

44. The nucleic acid construct of claim 43, wherein said intron is intron 3 of the porcine Gal α (1,3) galactosyl transferase gene.
45. The nucleic acid construct of claim 41, wherein said first construct sequence and said second construct sequence are homologous to different introns of porcine Gal α (1,3) galactosyl transferase gene.
5
46. The nucleic acid construct of claim 45, wherein said first construct sequence is homologous to an intron 3 region and said second construct sequence is homologous to an intron 4 region of porcine Gal α (1,3) galactosyl transferase.
47. A cell transfected with the nucleic acid construct of claim 33.
- 10 48. A cell transfected with the nucleic acid construct of claim 41.
49. A cell transfected with the nucleic acid construct of claim 44.
50. A cell transfected with the nucleic acid construct of claim 46.
51. A bacterial cell transformed with the nucleic acid construct of claim 33.
52. A bacterial cell transformed with the nucleic acid construct of claim 41.
- 15 53. A bacterial cell transformed with the nucleic acid construct of claim 44.
55. A bacterial cell transformed with the nucleic acid construct of claim 46.
56. A nucleotide sequence of intron 4 of the Gal α (1,3) galactosyl transferase gene having nucleotides 4938-11716 in Figure 1.

57. A nucleotide sequence of intron 5 of the Gal α (1,3) galactosyl transferase gene having nucleotides 11753-13748 in Figure 1.

58. A nucleotide sequence of intron 6 of the Gal α (1,3) galactosyl transferase gene having nucleotides 13810-14358 in Figure 1.

5 59. A nucleotide sequence of intron 7 of the Gal α (1,3) galactosyl transferase gene having nucleotides 14463-21627 in Figure 1.

60. A nucleotide sequence of intron 8 of the Gal α (1,3) galactosyl transferase gene having nucleotides 21766-27048 in Figure 1.

61. A lambda phage clone derived from a porcine genomic library comprising at least a portion of 10 the Gal α (1,3) galactosyl transferase gene, wherein the lambda phage clone is selected from the group consisting of pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2.

62. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 33, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating said embryo into an 15 appropriate host mammal, and generating offspring from said impregnated host mammal.

63. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 44, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating said embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.

64. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 46, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.

5 65. A transgenic mammal made according to the method of claim 62.

66. A transgenic mammal made according to the method of claim 63.

67. A transgenic mammal made according to the method of claim 64.

68. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 32, selecting for transfected cells comprising the nucleic acid of the 10 construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

69. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 44, selecting for transfected cells comprising the nucleic acid of the 15 construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

20 70. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 46, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate

host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

71. The nucleic acid construct of claim 43, further comprising a nucleic acid sequence
5 encoding a gene which is toxic to said eukaryotic cell

72. The nucleic acid construct of claim 71, wherein said gene which is toxic to said eukaryotic cell is the ricin A toxin gene.

1/27

Figure 1

Sequence from within Intron 3 to the end of intron 8

```

1 gtcgactcta ggcctcaactg gcctaatacg actcactata gggagctcga ggatcaatta
 61 gaggtccacc atcccttcc tgaatgccta aggccagata tggtaattt tagaattttt
121 caaatgcaga atatttatcc tatattatgt aacgcggccca gtgcagcaac agcacataac
181 aatgtacatc aatatttatg caaagaaaatg tttaaacagt ctcactaagt gataaagatt
241 ataaatagcc tcgtgtcagg gcctggatgc caactgaatt ataaacaggc ttttggttt
301 cagagcttgg agttggatga gggctgaga aactgctcca tggtaattt taggagggtt
361 gtgggtgtct ccagacccca cctccctccc aagctcttc accacccaca cttctctggg
421 agtgaagaca acggcagaga ggcattggcca cagtggccac agtctccacc cggatctgtc
481 tgctcccaaa cccaggcctt tcctcgcact cagtgcata tggtaattt taggagggtt
541 gtggctttt ccagcatctg ggcggagctg catgtagccc cgtgcatttc gtaactttgc
601 cctggggcccc gggctgttt tgccaggacc tgaggtgctc acaggaaaga agctccatct
661 ccccatcttc tcaccatctc tggAACACCA cctatcatga ttgtatctga aagggtggcga
721 ttgaatcaga ggtttccaaa cagagctcac tttccaaAGCA agaaggaaata gaggatcgat
781 gctgataatc ccatacttcc tcttctttaa ctggatttca caacagaggt gatggagcac
841 ctgagatcta agcctggagt cacctcgaaa ccctctctgc aaatattttt agaataaccc
901 gtcccctgaa aggacacatc tcagtgccat ttcatttca ttcacacatc tttttttttt
961 tttttttttt ttttttttgg gtttttggcc atttcttggg ccagtcctgc ggcataatgg
1021 ggttcccagg ctaagggtct aattggagcc gtagctgcag gcctacgcca gagccagagc
1081 cacacgggat ctgagcccgcg tctgcacact acaccacagc tcacggcaac gccggatctt
1141 taagccactg agcaaggcca gggatggaaac ccacaacccatc atgtttctca gtcagattcg
1201 ttaaccacac agccacaacg ggaactccca cacattatattt attgacggcc ttctctgtc
1261 tctgtggggc actggaaatt caggggtgat caagaagtca tccctcctgc cctcaggaag
1321 ctcaaaccac tcattattta ttgacggcc tctctgtct ctgtggggca ctgggaattt
1381 aggggtgacg aagaagtcat ccctcctgc ctcaggaagc tcaaaacaagc aggttagagga
1441 ggcagagcaa aatgcaggc ttatccggg agccgactcc cagggcgatg tgcacagcaa
1501 aggaatagag ggatggggc cggaggagag aaaagggtt cagccgtgtt caggggtgggg
1561 gtggaaagtg gtttcacaaa ggcagtgaca ttggctccca ggtgtccact ttctctgtc
1621 tgctaccttc tggctcttc ctctggcc ctccctatc ctacctctaa agcttcaccc
1681 acatcctcct ttcccttct ctctctggat tctctctgg gtaatcaaattt tcgttccctt
1741 cacgtcagat ccgatacgtt cttggtcca tgaacaactt ctccgattgc atggtctgcc
1801 tacatctctc tggatggat tagacttggaa tggccactt tctccctgtc cccttttagg
1861 tattcgcaca ctccccgaca ttacacacgtc caaaaggaa ttcatgatta ttatcctcca
1921 agctgttcc tcctccagcc catctgagaa aataactacaa ccccccgtct taagcagaaaa
1981 tcttgggtct tccttgcattt atctctgata acaaaaattac caaccacgtc ctatcaattt
2041 tctctccaaa gtatatatat atatatattt tttaattttt ttcccgctgt acagcatggg
2101 gatcaagtttta ttcttacatg tatattttcc ccccccctt tggatggatgg caatatgagt
2161 atcttagacat agttctcaat gtcactcagc aggtatctt tggatggatgg agttgtatct
2221 gataacccca agctcccgat ccctccactt ccctccctt cctgtcgggc agccacaagt
2281 ctattctcca agtccatgtt ttctttctt gtggagatgg tcatttgc tggatatttag
2341 attccagtttta taagtgtat catatggat ttgtcaaaatg atatattttt tttttttttt
2401 tcttttgc ttttgcattt ttttgcattt ttgtcaaaatg tggatggatgg tggatggatgg
2461 tacttggggc gtcggccgg catatggagg ccccccctt tggatggatgg tcggagctgt
2521 agccacccggc ctacgccaaga gccacagcaaa cggggatcc gagccgcgtc tgcaacctac
2581 accacagctc acggcaacgc tggatccatc acccaactgag caagggcagg gaccgaaccc
2641 gcaacctcat gtttgcattt cggatgcattt tggatggatgg tggatggatgg tggatggatgg
2701 gtatattttt aatcaaggcca cccttgcattt cggccacactt cctctttatg gtcatgagaa
2761 cggctgtccc ttgtccattt ctccatttc cacaactcagc acccaactgatgg gtctctctag
2821 gtggatggatgg atcaggggat tggatggatgg tggatggatgg tggatggatgg ccaccctact
2881 ttccatcatctt ttccagggtt tggatggatgg tggatggatgg tggatggatgg agcaccagcc
2941 actcaagggg gacagtgtca gtcactattt ttttgcattt cggatggatgg tggatggatgg
3001 tttcaaaaca cgagcagtc ttctctgtt ctgcctggta tagatgcgtt ttccaaaata
3061 ttctcatccc ttctcagggc ctttgcattt cttccatc ctatcttcat cccttggaa

```

2/27

3121 gctctaaagt catctccca aattgaaggg tgactaaaga gtttcccaga agaaaaaact
 3181 gagttccaa ctactacact gacttgcag aaatgttgt gtcttcatta aatgaaaaag
 3241 aaaaaactgt aacaagatata gaaaaataac agaaaaggaaa taataagact agaaaagtca
 3301 aatataatgt gaaggtttg catcaaacac ttaaataaac tagtacagat gttaaaagac
 3361 taaattatata agttgaagga tagctgtaa gatgtaaact atgacatcta aaacacaaaa
 3421 tggcggtt cccgtcacgg cacagtggaa acgaatccga ctaggaacca tgagggtgca
 3481 ggttcaattc ctgccttgc tcagtggtt aaggatccgg tggtccgtg agctgtggtg
 3541 tagtagcca atgaggctt gatcccgcgt tgctgtggct ctgggtgtagg cccgtggcta
 3601 cagctccgat tcgaccctta gcctggaaac ctccatatgc cgcgggagcg gccccttaaa
 3661 aagacaaaaag accaaaaaaa aaaaaaaaaca aaaaaccac aaaaatgtgg gaatcagtcc
 3721 tctactatgt ttatgttatt gtcaagttt cctttatgt ctgtaatat ttgcgttcta
 3781 gatgttaggt ctctgatatac gtgtgcataat atgttaacca atgttatgtc ttccctctgg
 3841 attgatccct ttgttattat gtaatgccct actttatctt ttgttacatt ctttgtttat
 3901 gagtattgtct gatatgtggc tagtgcac acctttcttgc tccttccat ttacaataaaa
 3961 tatctttcta tctccacca aattaaagta ctccgcaacc tgttattcca cccagcatcc
 4021 ctcccttcaactacaat ttcatgcac gatcaagaaa tagaatgtac cgactgttt
 4081 ccattgggtgg ggcatggaa aagtgggtgg aaagtgcaga gcttagatta taaaggccag
 4141 ggtgagagtt cccattgtgg tcagctgaaa tgaatctgac tagcatccat gagcacgaaag
 4201 gtttgatccc tggcctcaat cagtggttta aggtctggc gttgctgtcc gtgagttgt
 4261 gtgtagttcg cagacaaggc gtggacttag tggactgtg gctgtggcat aggctagtg
 4321 ctacagctct gattgcaccc ctggctggg aatctctata tgctgtgagt gtggccctaa
 4381 aatttaaatg aaattaaata aaggaccagg gtatattttt ctttgaggat aaggtacata
 4441 gtcagtatata caaggacagt agaccttaga aacggatgtc tcctctagtc tggatgcga
 4501 ggtggggcat ctgagttggg ggcggctgaa gccccttaggg accattaact aaaccgtca
 4561 ctctcccaca tctcggtgaa ccttggatc agtcaggatg ctcccccattt gagcctcaaa
 4621 atggccttag tattttttttt aaccacacg gcccctgttag ttcattgtact tggctaattt
 4681 gccagtgttag gcctatgca attaaggttag aacgcactcc tttagcgtcg ttgactattc
 4741 atcaactttt ctttttagaa aagatattgg tataagact tcttaaaaaa ccatattcca
 4801 ctctgggtgt atttaatcta atttccctt ctccctttt tttcccagga g-----

-----EXON 4-----

4938 -----gta attatgaaac atgatgaaat gatgttgatg aaagtctccct
 4981 ctaatctccat agttatcagc caagtcacca gcttcattaa aagtaggat tcactgacac
 5041 cgtaaagaaa gcattccaga gagttgcgt tggctcgat gggcagcaaa cccaaattagg
 5101 atccaagagg aggtgggtttt gatccctggc cttgctttt ggcttaagga tccggcattt
 5161 ccgtgacctg tgggttaggt tgcagatgc gctcgatctt ggcattgtg tggctgtggc
 5221 gtaggctgggt ggcttcagct ccagtttgac ccctagctg ggaacttcca tatcccacac
 5281 ttgcggccct aaaaatcaaa gaaagaaaaga aaatattcta cccttcctgt atccctgagc
 5341 ccttaataac cgtttttaaa gtcatttagat cttcaagtagc cttccagctt attaattatc
 5401 ttcccttcgt ccatgttgc attgtcctga tttttatacc tctgcagttc tgggttaggct
 5461 agagccagaa ataataaggt catgttaaga ccaagatata atattaaattt atttatatga
 5521 ccagatatgg aagttaccc tggactttc agacaggaat tccatgagaa atacaccctg
 5581 atttttgcaa tcctaaaata tttgcagatg taaaaggaac aactcaagtt gttgactttt
 5641 gctgcaaaac acactgagtc gctgggtgatt catttgcgtt tggctaaact tttgggtgtt
 5701 ttgtctttttt ttttttttaa ctctggaaag caaaatgaat taaacatttc tgagtttca
 5761 aattcatcag tggattcacc ccaaataattt gagctgcattc tttgttttg gaaactacga
 5821 tgccttggag attccagctg gagacgcattc tgacagaaag aaatgtctgc aagcagctac
 5881 aaaaatgcattt gatggcttgc acttaagagg cattgatacc gcttggccctt tctttcaaaa
 5941 agggccaccc tcaacttggc ctgaaggcat tcccgtgtg gtgcagcgaa aatgaatct
 6001 gacttaggaac cccgagggttggggtcaat ccctggccctt gctcagtgcc ttaaggatcg
 6061 ggtgttgaag taagctgtgg tggatattgc agacgcagct tggatctgg tggctgtgg
 6121 cttgggtgtt ggcggccagc tacagctcca cttggacccc tagtctggga accttttagg
 6181 tggcccttaa aaaggaaaaaa agacaacaaa caaacaacaaa accaaaaaaac aacttggcct
 6241 ggagagctat gtcatcacca ttgtatattt gatgggttagt gttttagtag cccctcaagt
 6301 tcaggatgtt ggcctggatt aacattagaa tggatctttaa attctacgac ttgatgagcc
 6361 agcaggacca ttttggccac ttagaaagga actgcattt caggtccatc agttagaagga
 6421 ggattctcta gggagttctc tcttagctca ggggttcaa gaattcagtc ttgtccctac

3/27

4/27

9901 ccataggtgg taggaaggca tgtatttcat tcctaccagg agatgtggac tcccagctgg
 9961 ggcagaaggc agagggagga gatcggggct ttggcagaat ctcaaacaaa tattagtgg
 10021 tagtggttt ttgttctgt tttaagagat gagggcagggc gtttccgatg tggcgcagtg
 10081 gaaacgaatc tgactagat ccatgaggat gcaggttca tccctggact cactcagtga
 10141 gttaaggatc cggcattgcc gtgagttgt gtgtnggtca cagacacacgc tcagatctgg
 10201 cattgtgtg gctgtgggt aggctagcag ctgtagctcc aattcaaccc ctagcctgg
 10261 aacttccaca tgccgcaggc gcaacccaa aagataaattg aataaataaa taaatatgcg
 10321 accttccccc cttggggctt ttgcatttt ttctctgt taggcacact cttgctaattc
 10381 ccttcaact gggcctccta agtacccctc agaactcagc taaaacatca tccccctcccc
 10441 tggggagcc tggagggttt cctgttaagt gtcctatgc ttcttggag ttttgaagtc
 10501 ctataatgtat gtgtttatca aaatagggtc caccctccct gccagttctt ctacaccaca
 10561 gacacatgtt gtctgtttca gtcaacactg tatgtctggc acttgacatg taacgcattgc
 10621 tcagcaggta ttgttgaat gaatggagggc ggtctgttag agtcgtcata tatttactga
 10681 tcccgcttgc taggatggc tcactgttt tgtagctta agaagtaccc ttttttttt
 10741 ttttttaat ggccacaccc atggcatata gaaattccac gaaggaagga agaaagaaag
 10801 aaagaaagaa gggaaattctt gggcaggga ttgaatccaa gccacaggtg caacctgagc
 10861 tgcagggttgc gcaacaccac atccttaac ccactgtgtc gggccaggga tcataccctg
 10921 gcatgtacag cgacccaaagc cacggcagtc agattcttt tcttccttc ttcttctt
 10981 ttcttttctt tttttttttt ggcttttgc cttttctagg tgccggcatat
 11041 ggagggttccc aggcttaggt tcgaatcaga gctgttagacg ccggcctaaa ccacagccac
 11101 agcaacacag gatccaagcc ttgtctgtca cctacaccac agtcacccgc aacgctgat
 11161 ccttaaccggc ctgagcggg ccaggattt aacccgcaac ctcatggc tttagttgat
 11221 tcgttaacca ctgagccatg atggaaactc ctgcagtcattttaacc cactatgcca
 11281 cagcaggaac tcctagaatg gccccttgag gctactctgt agacagctt gaggccagcg
 11341 ggcaagaccc ttttttctgg aggaagatata atcctgggtt agggatgggt gggctgtgg
 11401 cttcctggga cccatcttc gaggctctt ccctcagcaa agccacccgt gacaataaga
 11461 gctgccatct atttttttt ctttaacta agatttata ttttccagag acctccctc
 11521 ccaccgtttt atctgatcaa ttctgaaatg acgagatcc cgtgatatac tttttcgat
 11581 ctgcagggtt gaaacccctgg agtagccaca acccaggctc tcagctcagc cttagggtt
 11641 aatgataatg attgcaaaat agcttttctc tgatttccaa gtaacatgat atgttttt
 11701 ttccatttgc ttttag---- ----- **EXON 5-----** ----- --gtaaatgc
 11761 ttgttaattcc aaatatctt aggtcaccc ccatgtgacc ctggggccc tacagtccat
 11821 tcttaacatg gcagggtggc accgactgtt ggtcttagt ggaggagagg gatgggggtc
 11881 caggggtctg agctgtactt ctcagcccc tagacttgc ttttagagc atgagttgt
 11941 ttttccctt gcttctcatc aagtatctat ctcttaatg gatgttggg ggagaacatt
 12001 cctgccttgc tcataaaaaaa gaatcagat agatattatc cattatgcta cctactacat
 12061 gtgttataaa gacccttgc agaaattttt ccaagacaaa ggatttagaa gaaaggctgg
 12121 gtgtcctgtt aaactgtgt gtgttattt attatttattt attattacta ttactggta
 12181 ttaaggactc taagcctca tttttttttt ttttttttctatcttc cttgggttgc
 12241 aggaagctt gacaaagta ttgtgtttaa atgcttgcatttcttgc cttcatttt
 12301 tttaaaacat ttttcttat taaagtatag ctgatttata gtgcatttca tctgatata
 12361 ttatccctt ggtgttaat cctggctttt gtttagatgc atggatctt ggcaatttgc
 12421 tcaaaactcat tttgccaata tcttagctat gaagttaaaa taaagttaaa gatttttttc
 12481 tcacagagtg gctgggatga cccaaatgtt gtggaaacac ccgagtgact aaaatgtt
 12541 tctgtttctt tttgtttttt tttgattttt gtatgtttt cctatttata gtaaccacac
 12601 ttcttcata agccatttca agcaatttcc gaaagtatgat ggactttaag ttttttgac
 12661 ttccagttgt ggtcagttgc aaacaaatct gactgtatc catgaggatg catcttcgat
 12721 ccctggccctt gctcagttggg ttaaggatct ggtctgtctg tgacctgtgg ttaggttgc
 12781 agaggcggct cagattccaa gttgtgtgg ctgtggcata ggccggcage tacagctcca
 12841 attagacccc tagcctggga acctccatat gcccgggtg cggccctaaa aaaacaaaaa
 12901 aaaaaaaaaa aataataaaa taaaataaaag taagtttttctt actactatgc ctattccctgc
 12961 ctcaccaccc tatttgcata accaaaggc atggcagtc ctggctttcc cccttcttgc
 13021 cctcctctag acggtccacg ccttggagag actctcacag tgagtgtggc catctgagca
 13081 tttgtgtggaa gaaaaacggc ttctgattac agttgtgtca gcttcgggtt tagaaaccc
 13141 cagtttccctt gaaatctgtca ctgtgtaccc ttgttagcaac tccagttctc ctcatttagc
 13201 aagagatcat tcttggaaat gtgggctttt ctttttaag actacaactt actacagtaa
 13261 tttcagttt ataaagctt ttacgtagg ctactttttt tagtcttcga agccatccat

5/27

13321 cttcatggta ggtttgggtgc tgcattgtt ctctaactt ccagaggagg caatgggtat
 13381 tcagagagga aaaatgtctt atcagggtta cacaactgga aaggggcaga tccaggactt
 13441 ggatttatgg attagtgcctt tccccctgtt cattttttt ctccttctt ctcctgcga
 13501 gggctccctg gaccagctga gtgcgggggg ccacccaagg gaggggagat ctggggagtg
 13561 tcgggtccat cctctgtcat tcagctggag cagttcatt tctgttcatt ttacgaattt
 13621 gtgattttttt aaaaatgtat tttaagtgtat ctggccataaa aagagtttgg aacacccccc
 13681 accccctgccc ccagtttggg actgcctcat tttctaaaaa atgtttgtca tctttttcat
 13741 ttcttag--- ----- exon 6-----
 13810 ----- gtaagactgg gaaacggca tctgtgtatc tgctcaaggc tgtagagttcc
 13861 aaataaaaatg gtttcacagc catgacccctc tcagtcgcct cgtccttctg gcatatttgg
 13921 atcatgcaag gactctgtt cgtgggtgggg atgggcattt agttttttt tcaagggtttt
 13981 gtttttccc atttgatgtt ggagttagtg aggctcaggc tgaagttgag gctgcaggaa
 14041 aggctgtcat gagcttacat acgctgataa ggactgtttt atcagggttc cctgctcctg
 14101 acagaacctg agaattgtt ggtcattaca ttgcttaactg tgctggctg gaaggatgat
 14161 gaaaaagggg agcactgaac ctttcatca gacggacgct tttacgaata aatgttttagt
 14221 atcacaagcc ttctgttagac atttgcgtt cgcacgtgtt ctctgcctt gaatctgttt
 14281 tcacaagatc ctcaactctca ttctggcaca tgggtcacc cccctgcctt cctcagctt
 14341 ttttccgtt gtacgttag--- EXON 7-----
14463 ---gtaagaa aagaagcggtt gcccttatttc agtaaatcca cagtcagca ggaggtggca
 14521 gggctgtgtt aggcaacagg gccatgagaa cattcccgac agggagaca gagtttttagc
 14581 atctattttaga cccttaaaac aggtccggc ttttgcgtgt gtgtgtgtgt gtgtgtgagt
 14641 gactggcgtg gcatccccc aggtggggag tctggatgtt ttgactcaaa tggtgcgc
 14701 acaacatagt gattctgaca gcaatttagcc aatggcaaca gatgaatcac aaaataactgg
 14761 ggcttagcagt tcccgtgtt gcacagcgga aatgagtcca agtaggaacc gtgagggttc
 14821 acgttggatc cctggcctca ctcagtgggc taaggatccc gcttgcgc gacgggtgg
 14881 gtaggttgcg gacgtggcgc ggatctgggt ttgctgtgtt aggccggcag ctgtagctcc
 14941 agtttgcacc ctagcctggg aacccctcata tgctgcgggt gcagccctaa aaagcaaaaa
 15001 aatactgggg ctagggaaagg tcctatagcc catgggggtgc tggaaagc cacagaatct
 15061 agaatcagat aaagaaaatt cttaggttgc gtttaactctt ctgtaaaatg gcaatgacgc
 15121 gagttcccat tgggttcag tggaaaacaaa cccaaactagt atccattcta tccccagccc
 15181 cactcagtgg gtttaaggatc tggcattgtt gcaagctgtt gtgttaggtca aagacaaggc
 15241 tcagatctgg tggctgtgt gctgtgtgt agactggca ttacagctcc gattcaaccc
 15301 ctagactggg aacttccata ggctgcaccc tggccctaa aaaaataaaa ataaaataaaa
 15361 atggcagtga tgatttaaaa aacccatcaa ggccatgcac acactctaag caaacataca
 15421 acaatgttca taaaacacc tatgatatgc tagcccatgt gtgtcagact atcattttagt
 15481 ccattaatgt gtttagaaac tggcttaagg gtcaaaaatt gaaaccaaatt aggaaatcag
 15541 aaggcatctc atgaagtaag gataagtatt gtttagtggc attttgctt ccatatgtgt
 15601 gtgtgcgtgt gtttgcgtt taggatataat ttcaactatga gtcacgggtac tagacattgt
 15661 gtaggcaca ggagacagac aggagaaatt aatcaataca cttatattca agaagtttac
 15721 agtcctatag gggcactaga tggaaacta gcatttgcgc agattggtaa ggcgggtgt
 15781 ttaacctgac tggaaagaat gtggaaaggga gtgaggaggt gggggctgtc cgggatgact
 15841 gggaggcctg agagactggt gcatgtgtt gccgttgacc agggaaaggga ctggaggagg
 15901 agatacaggt ctaggaatag gggaggtat gagcttgatt aaggatgttag gtggatttc
 15961 gtacctgtgg ggcatttggg tagatgtcat ggagagccta tcttagaccg tcagactact
 16021 gtaacagaat accatagacg ggggtggctt taaaacaacaa aaatgttattt tgcacgttc
 16081 tggaggctgg gaagtccaaat atcaaggccac cagcagatgt ggtgtctgtt gagaacccac
 16141 tttctagttt atacatgccc gtcttctcat tgggttcttta ggccatggaa gggggagaggg
 16201 agctctccag ggcttctttt ctaaggccac ccatccctt cctgagggtt ccacccctcag
 16261 gatctaatca ccccaatgcc ccacccctt aggttgggtt tcaacacacg aatttgggg
 16321 gggacataat cgtcagtgtt cagcaggccc tttaagctgg cagaagatat gtggtagagaa
 16381 catttggaaaga ggagccatca gagttcaggc cgaggagtag agccaagtct tatctgacat
 16441 cagcaactaa agggtcaggg gagggaggaa aagaagtctg aaacggagtg ggagcagcag
 16501 tagaggttagg aggaggatcc gagggagaata gcatctattc cagcatgagg ggtcttggac
 16561 tcttggaaagg ctgagttatgt tcttggcaagt ttcttactgt ttttaggtgg ggaggtccat
 16621 gtcaagtgtt aaagaatggg aattttggcc tccccatgtt ttatttgagg agctggcact
 16681 aggttagcca tctatggaaat gctaagctac ttcttggaaag tggaaagaatg tggaaaacgc
 16741 tatacctttt tttttcttt tgagggggca cacctgcacg atatggaaatg tccctggc

6/27

7/27

20221 cggtcataacc tggggtgtga aggcaagagct ggggtgtggg gttttagcag cctagccatg
 20281 tccccctccat ctctgcagaa ttactcttag gggaggaagg aaggaggctt gaagcaggcgc
 20341 atgggggtggg ggtggggaggg cagaggccct ccagaatacg ggtcagagc ttgggtacgca
 20401 gggacagaag gcatgagcaa gaagtggggg tgacccagcc ttgcccacc tcgcaggaa
 20461 acatggtttc tccccatgggg gaggggcagg tttgtaaaaa cccttcctcc acccgcgtat
 20521 gaggggtgggg gtatgttcc gctgggctaa catctcttag acagcctcac cagggccata
 20581 cccaggcagc cagcaggggc ccgatgggac ccccaagacg ttgccccaaac cactggcttc
 20641 tacctgcagt ggctgaaata caagcacaca gctagcaggg gccaggccct ctcatgttgt
 20701 gctgggggtt gaatctaaat gaccacgtt cagttctttt ccgacgggaa agtgcagcca
 20761 cttccagagc aaatggggtt tgagccaagt ctctctagc ccctctgccc tctggggggc
 20821 cccagctgt acccagactt cccgcattt ctcttacaa ttttgcgtt gctcttcgt
 20881 ggcttcgtt gtttgcgtt gctggcgtt gtttgcgtt agttgcgtt gtcagagaa
 20941 catgccccctt cccctcccgcc ttttgcgtt ctgtcraga ggcacatga tctgagagag
 21001 gtctgagccc gacccctgtg tgaacccat ggccacagtt agcccgccgg gtccttcctt
 21061 ttctgggtga ctgtctcctt ctgcgtaaaa tgaaccacaa caccagggtcc ctttagtctg
 21121 gggccctttt tgcgtgatccaa ccagggtccct acccagaaat tgcttcgtt cagaaggaa
 21181 actgagacag cttcttcctc tttcagctca aatgtcactc tctcatcccg ctgtcaagc
 21241 catagggctt tctcagggtc aggtggcacc cactggata agtaacaccc aaagatgtcg
 21301 ctggcagctt aggaaggccc tggggagata ggcaaggggg ttttggagga agattggagga
 21361 cgaggacagc agaacaggccc gacggaaaggat acatgcattt tgacaggta cgatccccaa
 21421 aggggccacc agggcagccc tcagaggcac ctggccaga gcctcgtc cttcccccag
 21481 aagatgctgc aatgtcacaac caccagctga ctggggctaa aatacagtca ggattcaagg
 21541 ccagtccacca caagccatga ctgaccatgt ttcccccaga ctgtcgtaacc ttagcaaagc
 21601 catcctgact ctatgtttt tcaccagg --- EXON 8 --- -----
 ----- ----- ----- ----- ----- ----- -----
 21781 aaaactggcc ttgagttttt ccccttccac tatcagagga tgggtgggg gcccctgggt
 21841 ttacagagggc tggtcatgtc atgtctgaat tagtggagag gagaatgggt tcacaggccc
 21901 atttttagact cccttctgtc gaggccccca aaggctaaatg ataaaaacttag tcagagggtc
 21961 aactcttcc caccctcggt gaggggctt ggttgcaggg aagaagatct gctataccca
 22021 ctgcaccccg agtgcgtcga cagttacaccc acagccaccc ctggccctgac ctccacggcc
 22081 ctctgtggaa attccaaaaa tggcaattgt tagatggct gtgtgtgcgt cgtttctct
 22141 atgtctcgtc gaccccgac caaggacca agacagaatg gtcctaagtg gagtggttt
 22201 ccacgtccgc atggctgacc actcctctgt gcctctgccc tattcttcctt ggagggttt
 22261 gcccggggag attaggctgc tcagaactt cttccaga ggttggatag gttcctgtt
 22321 cagccccccgt ttccctgtt ggaactttt tcccaact gtaaacattt ttacttaaaa
 22381 gtagtagctg taatgttgcgt taaaaatata acccagttt cttttttaga gaatcccccc
 22441 ccctttttt caaaaacaaa agaaaaagtt caattttctt gttcacctcc gtgccccctc
 22501 cttcccccac atccatgggc cttccatct gtacctttc tgaaagccac aaagaaactg
 22561 aatcactttt ttatacagaa aaaattatca caacattgtt agacaactat acttcaacaa
 22621 aactaaaaaa aaaaatcaga aagagaaaga aaaaaaaga aaaaaggaa ggaaggaagg
 22681 gagttcccat catagctcat tggtaatga atctgacttag catccatgag gacacagggt
 22741 caatccctgg cttgtatcg tggtaagg atctgggtt gccgtggct gtgggttagg
 22801 tcacagatgc atttcgaatc ccgcattgtc gtggctctga cgcaggccat catcacagct
 22861 tcgattggac ccctagcgt ggagggtgc gcccctaaaaa agcaggaaag gaaggaagga
 22921 aggaagaaag aaaaaggaaagg aagggaaaga gaaagaaaaa gagagaaaaa gagaaaggaa
 22981 gaaaggaaga aagggaaaaa gaaagggaaaga aaaaaagaaa gaaagggaaaga aaaaaagaaaa
 23041 gaaggaagga agggaaaaag gaagggaaaga agggaaaaag gaagggaaaga aaaaaagaaaa
 23101 gaaagaaaga aaaaaagaaa gaaagggaaaga aaaaaagaaa gaaagaaaga aaaaaagaaaa
 23161 gaaagaaaga aaaaaagaaa aaaaaagaaa aaaaaagttt taggagttt gaaagattcc cagctcac
 23221 agagaagtga aggtcaaggt tgatctgagt gaagattaa gcgtccctt gagaccttct
 23281 ccaacctaga gcccagaagg tattactgtc actgtactaa tcaagtcata gtgtccctga
 23341 atgtactgtat cccaggccc tgggtggcat ctctaataga gaaggggtac tctggagttt
 23401 tgacctttcg actagaagaa tatgttctgt taggagttt gaaagattcc cagctcac
 23461 aatccttgg tataagagga gaccagccag ccatactgca gattgattt accttgggt
 23521 ttccaaaaag aacggctcca cattggctt agccatttt ttgcagctt agagaactaa
 23581 actgtacttg gcatgtcctg gtggaaaccgg cgcaactctgc tctgaggcga gtgaggactg
 23641 ttgtctcag cagcaataaa tcctcccaag agtggattt gacttttagg aggtgtatg

8/27

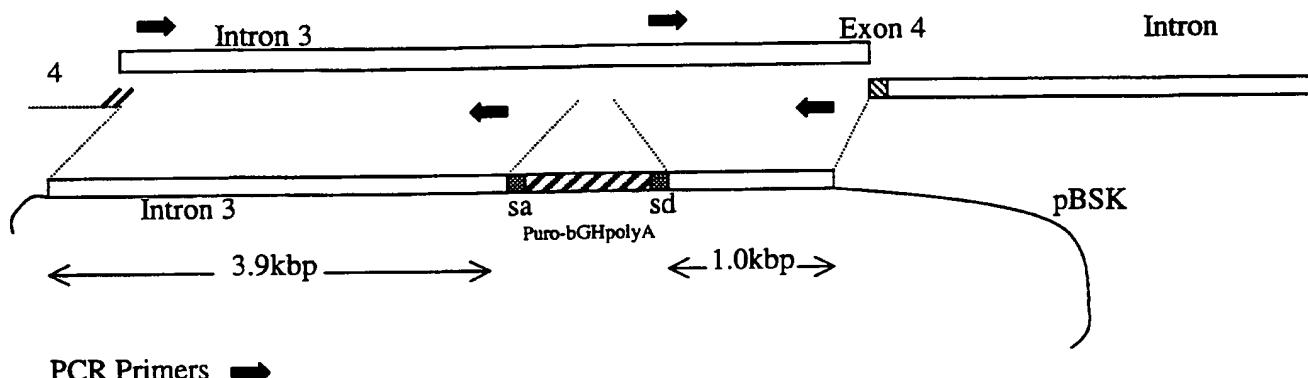
23701 tttggtttat tttcaaataa caagtgcaca gcaccaactg cagcgctta atctctaaac
 23761 aatgttaaac ctcacaagg tttctccct cgcccctagg atcacctccc aagcttcagg
 23821 gcacaataa tttttctta aagttccaga gtgcagaaga atttcctggc cattagcccc
 23881 agagccatgt tgaagtcct cttgtattt ctgagatatt ctctccgtt aaagaatcaa
 23941 tactataaaac ctggggttc cagagaccag cattgaaagc acagttatg tttgtatcaa
 24001 tcaaaaggaa tccaaggc aaccaaaaaa caaaaggaa tgccagtc tattacaaaa
 24061 tgtcatcatt ggctgataat gtatcaaca aattaaattt agtgcacagt cactggcat
 24121 gagtcgggt ggtgacgtgg gttgtttaa gccagccaa tacttcttc cttctctatc
 24181 atccccatgg cttcacagaa tagcctgtat ctggcttac agaagacagg aatgtatgg
 24241 gttttttttt ttctctttt cttttcagc tgcacctgtg gcgtatggaa gttctgagcc
 24301 acagctgtga cttctccat agtgcagca acactggatc tttaaaccac tttgtccactg
 24361 ccacaatggg aactcccccc tttcttttc ttccctgaaca tgatggttt tgaatgaata
 24421 gttaaatagc taaaatacca ctacacatct attagaatga cccaaatccc aaacactgac
 24481 aataccagct gctgatgaaa atgttagagcg atagggaaagc ttatttcactg ctgctggca
 24541 tgcgaaatga tacaaccaat ttgttaagaca gtttggcaga ttcttacaaa actaaacata
 24601 ctcttaccat acaatccagc gatggcacac ctttagtattt accctcaaaa gctgaaactt
 24661 ttatgtctac aaaaaaactt gcatacagat gttttagca ggtatttttt tttgggggg
 24721 gtgggtttt gttttttttt tttttttttt ttttggcca cacctgcggc atatggaggt
 24781 tcccaggcta agggttgaat cagagctaca gtcgcggcc tatgcccacag ccacagtaac
 24841 cccgagatctg agccacatct gcgcacctaca ccacagctca tggccacact ggatccttaa
 24901 cccctgagc gaggccaggg attgaacctg caactagtg ggtttgtga ccactgagcc
 24961 acgaaaggaa ctccgttat agcaacttta ttccataattt cccaaactt ggaagtaaca
 25021 ggtcttttag taggcaattt aattgataaa ttgttagcaca ttccagacagt ggagtattag
 25081 cagtgttta aaaaaaaaaa tcaatccatg aaaagacatg gaggAACCTC aatgttcgt
 25141 tacaagaaaa ccacatcgaa agactaccta ctatatgtt ccaactatata gacattctgg
 25201 aaaaaggtaa aggtactgtt ccatggagac agtaaaaata tcagtagatg ccaggcgatg
 25261 gtgtgggaga atggatgaat agacagaaca caggattttt tagggcagtg aaaataactct
 25321 gtatggtacc atacatagt atagataaaat gtcattttac atttggtcg attcatagaa
 25381 agtctgccac caagagtggaa ccttagtgtt aactacggc ttgggtgtat aatgtatgt
 25441 cagtgttagtt tcatcaggta caacacgtgc taccctggc ctggatgtt atgggtggag
 25501 aagctgggag tgggagttggg gggagcgggg ggaattctgc actttcagct caatttgtct
 25561 gtgagcctaa acatgttcta aaaactaaaaa tctatattttt taaatgaaac gagatagtt
 25621 aagtaagggtt gaatcagagg tcgttgc tcatatattttt atgggtttta taatgtatgg
 25681 gtcgtgataa tgggctggc tttgggtgc cagcgaaac tcaacaggta aatcagtaag
 25741 atggaagaat atccatattt cataatccat tgggggtgc ccagagaagg ataagagagt
 25801 ggcgatgggtt aggcttaatg accctgcac atcacgttgc catctggtag gtcagtaaaa
 25861 gtattggat ataggctggg gttttggagg aagagctaag aacttagtgc ccctgtatgg
 25921 aatcaaaggaa tacaacattt cagttataag atgaataa tctggatct aatgtcagca
 25981 atgggattgt tttatcat actgttaat atacttttaa gttggtcaga gaataagtcc
 26041 taagttttct tggccacaaaa aatgttaat tatgtatgat gatgaagggtg tcaagtaatg
 26101 ccacagtgtt aatcattttt aatatatgag tgcattttt gatgggtttt gtcactcaag
 26161 tgggttaaga atctgactgtt taaccatgag gatgggtt cgtatccctgg cttgtcttag
 26221 tgggttaaag gatcaagcat tgccaccgc tgccgtgtg gtcatagaca aagctcagat
 26281 ctgggtgtgc tgggtgtgc gcatggggca gcatgtgc gtcagtaattt accctgttc
 26341 tgggaacctc catatgtgc ggtatgtggcc cttttttttt aatgggggggggg
 26401 aagtgcatca catcaacact ttgcacaccc taaaacttaca caatgtatata tgcgttat
 26461 atctgaataa tgcaggaaaa aaggaaagac tagcatctga actgtattt aaggacaagg
 26521 ggataaaagac taattaaggc agatgttaat gagttatcac tttcattaca gttaaagctg
 26581 gtgtggagt atgtggctt taactgcagg ccctgggtt ccaaatctt aaccacagaa
 26641 ttagaaagac ttaggcatttcc aggcacatac agcacaagta actgcaggcc tctggatgt
 26701 ggagcaggcc ctgtgcactt aaggttaccac tgcaaggccc agagcagctg aaaacacatg
 26761 ttctctctgc ctgggtggct tccaaagatgtt agagaggaag gaggcaggct gaggcatgccc
 26821 agccaccctg ccagaatcac cagtcaggta agccactcca cttccccc aatgtatgac
 26881 tgaatggtgg agatgtgtt ggaatgttac agcaacagac gtctctcatc caggatgggg
 26941 aaaaatcatt ctttcctaa actgcataat acagactaga tgataatagc atattgttcc
 27001 ctctagaaat cccagaggat acatttaccc cattttttt tatttcag-- -----
 ----- --EXON 9

9/27

Figure 2

Promoter trap Construct: Insertion of Engineered exon into the intron of a gene.

Endogenous alpha-Gal transferase gene diagram



PCR Primers →

sa = splice acceptor

sd = splice donor

pBSK = pBluescript cloning vector

Puro-bGHpolyA = puromycin resistance with the bovine Growth hormone gene poly A acceptor signal sequence

10/27

Figure 3**A.**

10 20 30 40 50
gacggccgcAGGCCTCACTGGCCTAATACGACTCACTATAGGGAGCTGA
 CAGCTGAGATCCGGAGTGACCGGATTATGCTGAGTATCCCTCGAGCT
 60 70 80 90 100
 GGATCAATTAGAGGTCCACCATCCCTTCCTGAATGCCTAAGGCCAGATA
 CCTAGTTAACCTCCAGGTGGTAGGGAAAGGACTTACGGATTCCGGTCTAT
 110 120 130 140 150
 TGTTGGAATTAGAATTTCAAATGCAGAATATTATCCTATATTATGT
 ACAACCTTAAATCTTAAAAAGTTACGTCTATAAATAGGATATAATACA
 160 170 180 190 200
 AACGCCCCAGTGCAGCAACAGCACATAACAATGTACATCAATATTTATG
 TTGCGGGGGTCACGTCGTTGTCGTATTGTTACATGTAGTTAAATAC
 210 220 230 240 250
 CAAAGAAATGTTAACAGTCTCACTAAGTGATAAAGATTATAAATAGCC
 GTTTCTTACAAATTGTCAGAGTGATTCACTATTCTAATATTTATCGG
 260 270 280 290 300
 TCGTGTCAAGGCCCTGGATGCCAAGTGAAATTATAAACAGGCTTGGTTT
 AGCACAGTCCGGACCTACGGTGACTTAATATTGTCGAAAACCAAAA
 310 320 330 340 350
 CAGAGCTTGGAGTTGGATGAGGGTCTGAGAAACTGCTCCATGTTAGGGT
 GTCTCGAACCTAACCTACTCCAGACTCTTGACGAGGTACAAGTCCCA
 360 370 380 390 400
 TACCCAGTCTGGGTGCTCCAGACCCCCACCTCCTCCAAAGCTCTCTC
 ATGGGTCAAGACACCCACAGAGGTCTGGGTGGAGGAAGGGTCGAGAGAG
 410 420 430 440 450
 ACCACCCACACTTCTGGGAGTGAAAGACAACGGCAGAGAGGGCATGGCCA
 TGGTGGGTGTGAAGAGACCCACTTCTGTTGCCGTCTCCGTACCGGT
 460 470 480 490 500
 CAGTGGCCACAGTCTCCACCCGATCTGCTGCTCCAAACCCAGGCCTT
 GTCACCGGTGTCAGAGGTGGGCTAGACAGACGAGGGTTGGTCCGGAA
 510 520 530 540 550
 TCCTCGCACTCAGTGCTAATGCTGTTGATGTAGGAGTCAGTGCTTTT
 AGGAGCGTGAGTCACGATTACGACAACATCCTCAGTTCAACCGAAAAAA
 560 570 580 590 600
 CCAGCATCTGGGCCAGCTGCATGTAGCCCCGTGCATTCGTAACTTGC

7/1/97

GGTCGTAGACCCGGCTCGACGTACATCGGGGCACGTAAAGCATTGAAACG
 610 620 630 640 650
 CCTGGGCCCGGGCTGTTGTGCCAGGACCTGAGGTGCTCACAGGAAAGA
 GGACCCGGGGCCCGACAAACACGTCCTGGACTCCACAGAGTGTCCCTTCT
 660 670 680 690 700
 AGCTCCATCTCCCCATCTTCTCACCATCTCTGGAACACCACCTATCATGA
 TCGAGGTAGAGGGTAGAAAGAGTGGTAGAGACCTTGTGGATAGTACT
 710 720 730 740 750
 TTGTATCTGAAAGGTGGCGATTGAATCAGAGGTTCCAAACAGAGCTCAC
 AACATAGACTTCCACCGCTAACCTAGTCCTCAAAGGTTGTCTCGAGTG
 760 770 780 790 800
 TTTCCAAGCAAGAAGGAATAGAGTGACATGGCTGATAATCCCATACTTTC
 AAAGGTTCGTCTCCTTATCTCACTGTACCGACTATTAGGGTATGAAAG
 810 820 830 840 850
 TCTTCTTAACTGGATTTCACAACAGAGGTGATGGAGCACCTGAGATCTA
 AGAAGAAATTGACCTAAAGTGTGTCTCCACTACCTCGTGGACTCTAGAT
 860 870 880 890 900
 AGCCTGGAGTCACCTCAGAACCCCTCTCTGCAAATATTGGAGAATAACCC
 TCGGACCTCAGTGGAGTCTGGAGAGACGTTATAAACCTTTATTGGG
 910 920 930 940 950
 GTCCCCCTGAAAGGACACATCTCAGTGCCATTCTCATTTCATTCACACATC
 CAGGGGACTTCCCTGTGTAGAGTCACGGTAAGAGTAAAGTAAGTGTGTAG
 960 970 980 990 1000
 TTTTTTTTTTTTTTTTTTTGGGCTTTGCCATTCTGGG
 AAAAAAAAAAAAAACCCGAAAAACGGTAAAGAACCC
 1010 1020 1030 1040 1050
 CCAGTCCTGCGGCATATGGAGGTCCCAGGCTAAGGGTCTAATTGGAGCC
 GGTCAGGACGCCGTATAACCTCCAAGGGTCCGATTCCCAGATTAACCTCGG
 1060 1070 1080 1090 1100
 GTAGCTGCAGGCCCTACGCCAGAGGCCAGAGGCCACACGGGATCTGAGCCCG
 CATCGACGTCCGGATGCGGTCTCGGTGTGCCCTAGACTCGGCCGC
 1110 1120 1130 1140 1150
 TCTGCAACCTACACCACAGCTACGGCAACGCCGGATCCTAAGCCACTG
 AGACGTTGGATGTGGTGTGAGTGCCGTTGCGGCCTAGGAATTGGTGTGAC
 1160 1170 1180 1190 1200
 AGCAAGGCCAGGGATGGAACCCACAACCTCATGTTCTAGTCAGATTG
 TCGTTCCGGTCCCTACCTTGGGTGTTGGAGTACAAAGGATCAGTCTAAC
 1210 1220 1230 1240 1250
 TTAACCACAGAGGCCACAACGGGAACCTCCCACACATTATTGACGGGC
 AATTGGTGTCTCGGTGTTGCCCTGAGGGTGTGTAATAAAACTGCCGG

12/27

1260 1270 1280 1290 1300
 TTCTCTGCTCTGTGGGCCTGGAAATTCAAGGGGTGATCAAGAAGTCA
 AAGAGACGAGAGACACCCCGTACCCCTTAAGTCCCCACTAGTTCTTCAGT

 1310 1320 1330 1340 1350
 TCCCTCCTGCCCTCAGGAAGCTCAAACCACTCATTATTGACGGCCT
 AGGGAGGACGGGAGTCCTCGAGTTGGTGAGTAATAAAACTGCCGGA

 1360 1370 1380 1390 1400
 TCTCTGCTCTGTGGGCCTGGAAATTCAAGGGGTGACGAAGAAGTCAT
 AGAGACGAGAGACACCCCGTACCCCTTAAGTCCCCACTGCTTCTTCAGTA

 1410 1420 1430 1440 1450
 CCCTCCTGCCCTCAGGAAGCTCAAACAAGCAGGTAGAGGAGGCAGAGCAA
 GGGAGGACGGGAGTCCTCGAGTTGTTCGTCCATCTCCTCCGTCTCGTT

 1460 1470 1480 1490 1500
 AATGCAGGTCTTATCCGGTGAGCCGACTCCCAGGGCGATGTACAGCAA
 TTACGTCCAGAATAGGCCACTCGGCTGAGGGTCCCCTACACATGTCGTT

 1510 1520 1530 1540 1550
 AGGAATAGAGGGATGGGGCCGGAGGAGAGAAAAGGGCTTCAGCCGTGGT
 TCCTTATCTCCCTACCCCCGGCCTCCTCTTTCCCGAAGTCGGCACCA

 1560 1570 1580 1590 1600
 CAGGGTGGGGGTGGGAAGTGGCTTCACAAAGGCAGTGACATTGGCTCCCA
 GTCCCCACCCCCACCCCTCACCGAAGTGTTCGTCACTGTAACCGAGGGT

 1610 1620 1630 1640 1650
 GGTGTCCACTCTTCTGTCTCTGCTACCTCTGGTCCCTCCTCTGGGCC
 CCACAGGTGAGAAGACAGAGACGATGGAAGACCAGGAGAGGAAGACCCGG

 1660 1670 1680 1690 1700
 CTCCCTATCTTACCTCTAAAGCTTCACCCACATCCTCCTTCTTTCT
 GAGGAGATAGGATGGAGATTCGAAGTGGGTGTAGGAGGAAAGGAAAAGA

 1710 1720 1730 1740 1750
 CTCTCTGGATTCTCTCCTGGTAATCAAATTGTTCCCTTCACGTCAAGAT
 GAGAGACCTAAGAGAGGACCCATTAGTTAAGCAAGGGAAAGTGCAGTCTA

 1760 1770 1780 1790 1800
 CCGATACGTTCTTGGCATGAACAACTCTCCGATTGCATGGTCTGCC
 GGCTATGCAAGGAACCAGGTACTTGTGAAGAGGGCTAACGTACCAAGACGG

 1810 1820 1830 1840 1850
 TACATCTCTGATGAACTTTAGACTTGAATGTCCACTTGTCTCCCTGTC
 ATGTAGAGAGACTACTTGAAATCTGAACCTACAGGTGAAACAGAGGGACAG

 1860 1870 1880 1890 1900
 CCCTTTAGGTATTGACACTCCCGACATTCACACGTCCAAAGGGAA
 GGGAAAATCCATAAGCGTGTGAGGGCTGTAAGTGTGCAGGTTTCCCTT

 1910 1920 1930 1940 1950
 TTCATGATTATTATCCTCCAAGCCTGTTCTCCTCCAGCCCCATCTGAGAA

13/27

AAGTACTAATAATAGGAGGTTCGGACAAGGAGGAGGTGGGTAGACTCTT

1960 1970 1980 1990 2000
AATACTACAACCCCCCTGCTTAAGCAGAAATCTGGGTCTCCTGTCTC
TTATGATGTTGGGGGACGAATTCGTCTTAGAACCCAGAAGGAACAGAG

2010 2020 2030 2040 2050
ATCTCTGATAACAAAATTACCAACCACGTCTATCAATTCTCTCTCCAAA
TAGAGACTATTGTTTAATGGTTGGTGCAGGATAGTTAAGAGAGAGGTTT

2060 2070 2080 2090 2100
GTATATATATATATATTTTAATTTTCCCGCTGTACAGCATGGG
CATATATATATATATAAAGGGCGACATGTCGTACCC

2110 2120 2130 2140 2150
GATCAAGTTATTCTTACATGTATATTCCCCCACCCTTGTTCGTTG
CTAGTTCAATAAGAATGTACATATAAAAGGGGGTGGAAACAAGGCAAC

2160 2170 2180 2190 2200
CAATATGAGTATCTAGACATAGTCTCAATGCTACTCAGCAGGATCTCCT
GTTACTCATAGATCTGTATCAAGAGTTACGATGAGTCGTCTAGAGGA

2210 2220 2230 2240 2250
TGTAAATATAAGTTGTATCTGATAACCCCAAGCTCCGATCCCTCCCCACT
ACATTATATTCAACATAGACTATTGGGGTTCGAGGGCTAGGGAGGGTGA

2260 2270 2280 2290 2300
CCCTCCCTCTCCTGTCGGGCAGCCACAAGTCTATTCTCCAAGTCCATGAT
GGGAGGGAGAGGACAGCCGTCGGTGTTCAGATAAGAGGTTCAGGTACTA

2310 2320 2330 2340 2350
TTCTTTCTGTGGAGATGGTCATTGTGCTGGATATTAGATTCCAGTTA
AAAGAAAAGACACCTCTACCAAGTAAACACGACCTATAATCTAAGGTCAAT

2360 2370 2380 2390 2400
TAAGTGATATCATATGGTATTTGTCAAAGTATATATTATTCTTCTTG
ATTCACTATAGTATACCATAAACAGTTCATATATAAAATAAAAGAAC

2410 2420 2430 2440 2450
TCTTTTGCTTTGTCTTTTTGTTGTTGTTGTTGTTGTTGTTG
AGAAAAACAGAAAAGAGAAAAAACACAACAACAACAACAACAAAC

2460 2470 2480 2490 2500
TTGTTGCTATTACTTGGGCCGCTCCCGCGCATATGGAGGTTCCCAGGCT
AACAAACGATAATGAACCCGGCGAGGGCGCCGTATAACCTCAAGGGTCCGA

2510 2520 2530 2540 2550
AGGAGTTGAATCGGAGCTGTAGCCACCGGCTACGCCAGAGCCACAGCAA
TCCTCAACTTAGCCTCGACATCGGTGGCCGGATGCGGTCTCGGTGTCGTT

2560 2570 2580 2590 2600
CGCGGGATCCGAGCCCGTCTGCAACCTACACCACAGCTCACGGCAACGC
GCGCCCTAGGCTCGCGCAGACGTTGGATGTGGTGTGAGTGCCGTTGCG

14/27

2610 2620 2630 2640 2650
 TGGATCCTAACCCACTGAGCAAGGGCAGGGACCGAACCGCAACCTCAT
 ACCTAGGAATTGGGTGACTCGTCCCGTCCCTGGCTGGCGTTGGAGTA

2660 2670 2680 2690 2700
 GGTTCCTAGTCGGATTCGTTAACCACTGCGCCACGACGGAACTCCAAA
 CCAAGGATCAGCCTAACGCAATTGGTGACGCGGTGCTGCCCTGAGGGTT

2710 2720 2730 2740 2750
 GTATATTTGAATCAAGCCACCCTTGAGGCCAGGCCACCTCCTCTTATG
 CATATAAAACTTAGTCGGTGGGAAACTCGGTCCGGTGGAGGAGAAATAC

2760 2770 2780 2790 2800
 GTCATGAGAACGGTCTGCCCTTGTCTTCTCCATTCTCCACACTCAGC
 CAGTACTCTGCCAGACGGAACAGGAAAGAGGTAAGAGGTGTGAGTCG

2810 2820 2830 2840 2850
 ACCCAGATGGGTCTCTAGGTGAAGTTGGATCAGGGGATTCTCCAGCCT
 TGGGTCTACCCAGAGAGATCCACTTCAACCTAGTCCCCTAACGAGGTCGAA

2860 2870 2880 2890 2900
 TAGATGCTTTGGGATTCCCCACCCCTACTTCCATACCTTCCAGGTT
 ATCTACGAAAAACCTAACGGGTGGGATGAAAGGTATGGAAAGGTCCAAG

2910 2920 2930 2940 2950
 TGACTGCCTCTGCCCTCTGACTGCCTAGCACCCAGCCACTCAAGGGG
 ACTGACGGAGACGGGGGAAGACTGACGGATCGTGGTCGGTAGTTCCCC

2960 2970 2980 2990 3000
 GACAGTGTCACTATTCTTCTGTCCAGGTTTTGCTTTGT
 CTGTCACAGTCAGTGATAAAAAAGAACAGGTCCAAAAACGAAAACAAA

3010 3020 3030 3040 3050
 TTTCAAAACACGAGCAGCTTTCTCTGTCTGCCTGGTATAGATGCTGT
 AAAAGTTGTGCTCGAGAAAGAACAGACGGACCATACTACGACA

3060 3070 3080 3090 3100
 TTCCAAAATATTCTCATCCCTCTCACGGCCCTTGTCACTCCTTCCCATC
 AAGGTTTATAAGAGTAGGGAAGAGTGCAGGAAACAGTAGGAAAGGGTAG

3110 3120 3130 3140 3150
 CTATCTCATCCCTGGGAAGCTCTAAAGTCATCTCCCAAATTGAAGGG
 GATAGAAAGTAGGGAACCTTCGAGATTCAGTAGAGGGTTAACTTCCC

3160 3170 3180 3190 3200
 TGACTAAAGAGTTCCCAGAAGGAAAAACTGAGTTCCAACACTACACT
 ACTGATTCTCAAAGGGTCTCCTTTGACTCAAAGGTTGATGATGTGA

3210 3220 3230 3240 3250
 GACTTGCAAGAAAATGTTGTCTTCATTAAATGAAAAAGAAAAACTGT
 CTGAACGTTCTTACAAACACAGAAGTAATTACTTTCTTTTGACA

3260 3270 3280 3290 3300
 AACAAAGATATGAGAAAATACAGAAAGGAAATAATAAGACTAGAAAAGTCA

15/27

TTGTTCTATACTCTTTATGTCTTCCTTATTATTCTGATCTTCAGT

3310 3320 3330 3340 3350
 AATATATAGTGAAGGTGTCATCAAACACTAAATAACTAGTACAGAT
 TTATATATCACTTCCACAACTAGTTGTGAATTATTGATCATGTCTA

3360 3370 3380 3390 3400
 GTTAAAAGACTAAATTATAGTTGAAGGATAGCTGTGAAGATGTAAC
 CAATTCTGATTTAATATATCAACTTCCATCGACACTTCTACATTGA

3410 3420 3430 3440 3450
 ATGACATCTAAAACACAAAATGTTGGCGTCCCGTCACGGCACAGTGGAA
 TACTGTAGATTGTGTTACAACCGCAAGGGCAGTGCCGTGTCACCTT

3460 3470 3480 3490 3500
 ACGAATCCGACTAGGAACCATGAGGTTGCAGGTTCAATTCCCTGCCCTTGC
 TGCTTAGGCTGATCCTGGTACTCCAACGTCCAAGTTAAGGACGGAACG

3510 3520 3530 3540 3550
 TCAGTGGGTTAAGGATCCGGTGTGCCGTGAGCTGTTGTTAGGTAGCCA
 AGTCACCCAATTCCCTAGGCCACAACGGCACTCGACACCACATCCATCGGT

3560 3570 3580 3590 3600
 ATGAGGCTTGGATCCCGCGTTGCTGTGGCTCTGGTGTAGGCCGGTGGCTA
 TACTCCGAACCTAGGGCGAACGACACCGAGACCACATCCGGCCACCGAT

3610 3620 3630 3640 3650
 CAGCTCCGATTGACCCCTAGCCTGGAACCTCCATATGCCGGGGAGCG
 GTCGAGGCTAACGCTGGGATCGGACCCCTGGAGGTATACGGCGCCCTCGC

3660 3670 3680 3690 3700
 GGCCCTTAAAAAGACAAAGACCAAAAAAAAAAAACAAAAACCCAC
 CGGGGAATTTCTGTTCTGGTTTTTTTTTTGTTTTGGGTG

3710 3720 3730 3740 3750
 AAAATGTTGGGAATCAGCCTCTACTAGTATTATGTTATTGTCAAGTTT
 TTTTACAACCTTAGTCAGGAGATGATCATAATAACAGTTCAAAA

3760 3770 3780 3790 3800
 CCTTTTATGTCGTTAATATTGCGTTCTAGATGTAGGTGCTCTGATATC
 GGAAAATACAGACAATTATAACGCAAGATCTACATCCACGAGACTATAG

3810 3820 3830 3840 3850
 GTGTGCATATATGTTAACCAATGTTATGTCCTCCTGGTATTGATCCCT
 CACACGTATATACAATTGGTTACAATACAGAAGGAGACCATAACTAGGGA

3860 3870 3880 3890 3900
 TTGTTATTATGTAATGCCCTACTTTATCTTTGTTACATTCTTGTATT
 AACAAATAACATTACGGGATGAAATAGAAAACAATGTAAGAAACAAATA

3910 3920 3930 3940 3950
 GAGTATTGCTGATATGTGGCTAGCTGCCACACTTTCTGTCCCTTCCAT
 CTCATAACGACTATACACCGATCGACGGTGTGAAAGAACAGGAAAGGTA

16/27

3960 3970 3980 3990 4000
 TTACAATAAATATCTTCTATCTCCACCCAAATTAAAGTACTCCGCAACC
 AATGTTATTTAGAAAGATAGAGGTGGGTTAATTGAGGCAGCTGG
 4010 4020
 TGTTATTCCACCCAGCATCCgtcgac
ACAATAAGGTGGTCGTAGGcagctg

B.

Ligated to the following intron 4 splice acceptor sequence

Intron 4 splice acceptor sequence

11530 11540 11550
gtcgacCCACCGTTGATCTGAGTAATTCTGAAATG
cagctgGGTGGCAAAGTAGACTCATTAAGACTTAC
 11560 11570 11580 11590 11600
 ACGAGAGTCCCCTGATATCATTTTTCGATCTCGAAGGTGGAAACCTGGG
 TGCTCTCAGGGCACTATAGTAAAAAGCTAGAGCTTCCACCTTGGACCC
 11610 11620 11630 11640 11650
 AGTAGCCACAACCCAGGCTCTCAGCTCAGCCTAGGGTTCAATGATAATG
 TCATCGGTGTTGGTCCGAGAGTCGAGTCGGATCCCAAAGTTACTATTAC
 11660 11670 11680 11690 11700
 ATTGCAAAATAGCTTTCTGCATTCCAAGTAACATGATATGTTTTAT
 TAACGTTTATCGAAAAGAGACGTAAGGTCATTGACTATACAAAAATA
 11710
TTCCATTTGCTTTAGgaattc
AAGGTAAACGAAAATCcttaaag

C.

Ligated to Neomycin Resistance gene

gaattcAATGGATCCCCACCATGG-NEO-GGATATCCACTACTTAGTAATGCCgaagctt
CCTATAGGTGATGAATCATTATCGGCttcgaa

17/27

D.

Ligated to the intron 4 splice donor sequence

4940	4950			
<u>aagcttGTAATTATGAAAC</u>				
<u>ttcgaa</u> CATTAATACTTTG				
4960	4970	4980	4990	5000
ATGATGAAATGATGTTGATGAAAGTCTCCTCTAATCTCCTAGTTATCAGC				
TACTACTTACTACAACTACTTCAAGAGGAGATTAGAGGATCAATAGTCG				
5010	5020	5030	5040	5050
CAAGTCACCAGCTTGCATTAAGTAGGATTCACTGACACCGTAAAGAAA				
GTTCAAGTGGTCAACGTAATTTCATCCTAAGTGACTGTGGCATTCTT				
5060	5070	5080	5090	5100
GCATTCAGAGAGAGTTGCCGTTGGCTCAGGGGCAGCAAACCCAATTAGG				
CGTAAGGTCTCTCAACGGCAACACCGAGTCCCCGTCGTTGGGTTAATCC				
5110	5120	5130	5140	5150
ATCCAAGAGGAGGTGGGTTGATCCCTGGCCTGCTTTGGCTTAAGGA				
TAGGTTCTCCTCCACCCAACTAGGGACCGAACGAGAAACCGAATTCC				
5160	5170			
TCCGGCATTGCCGTGACCTGTGG <u>ctgcag</u>				
<u>AGGCCGTAACGGCACTGGACACCgacgtc</u>				

E.

Ligated to intron 3 3' sequence

<u>ctgcagCCCTCTTCAACTACAATTTCATGCAGC</u>				
GGGAGAAAGTGTGATGTTAAAGTACGTCG				
4060	4070	4080	4090	4100
GATCAAGAAATAGAATGTACCGACTGTTGCCATTGGTGGGGCATGGGAA				
CTAGTTCTTATCTTACATGGCTGACAAACGGTAACCACCCGTACCCCTT				
4110	4120	4130	4140	4150
AAGTGGGTGGAAAGTGCAGAGCTTAGATTATAAAGGCCAGGGTGAGAGTT				
TTCACCCACCTTCACGTCTCGAATCTAATATTCCGGTCCACTCTCAA				
4160	4170	4180	4190	4200
CCCATGTGGTCAGCTGAAATGAATCTGACTAGCATCCATGAGCACGAAG				
GGGTAACACCAGTCGACTTACTTAGACTGATCGTAGGTACTCGTGCTTC				
4210	4220	4230	4240	4250

18 / 27

GTTTGATCCCTGGCCTCAATCAGTGGGTTAAGGATCTGGCGTGCTGTCC
 CAAACTAGGGACCGGAGTTAGTCACCAATTCTAGACCGAACGACAGG

 4260 4270 4280 4290 4300
 GTGAGTTGTGGTGTAGTCGCAGACAAGGCCTGGACTTAGTGTGGCTGTG
 CACTCAACACCACATCAAGCGTCTGTTCCGCACCTGAATCACACCGACAC

 4310 4320 4330 4340 4350
 GCTGTGGCATAGGCTAGTGGCTACAGCTCTGATTGACCCCTAGCCTGGG
 CGACACCGTATCCGATCACCGATGTCGAGACTAAGCTGGGATCGGACCC

 4360 4370 4380 4390 4400
 AATCTCTATATGCTGTGAGTGTGCCCTAAAATTAAATGAAATTAAATA
 TTAGAGATATACGACACTCACACCGGGATTTAAATTACTTTAATTAT

 4410 4420 4430 4440 4450
 AAGGACCAGGGTATATTCTTGAGGATAAGGTACATAGTCAGTATAT
 TTCCTGGTCCCATAAAAAAGAAACTCCTATTCCATGTATCAGTCATATA

 4460 4470 4480 4490 4500
 CAAGGACAGTAGACCTAGGAAACGGATGCTCCTCTAGTCTGTGATGCGA
 GTTCCTGTCATCTGGATCCTTGCCCTACGAAGGAGATCAGACACTACGCT

 4510 4520 4530 4540 4550
 GGTGGGGCATCTGAGTTGGGGGGCTGGAGGCCCTAGGGACCATTAACT
 CCACCCCGTAGACTCAACCCCCGCCGACCTCGGAATCCCTGGTAATTGA

 4560 4570 4580 4590 4600
 AAACCCGTCACTCTCCCACATCTCGGTGGACCTGGGATCAGTCAGGATG
 TTTGGGCAGTGAGAGGGTGTAGAGCCACCTGGAACCCTAGTCAGTCCTAC

 4610 4620 4630 4640 4650
 CTTCCCTTGTAGCCTCAAAATGCCCTAGTATCCTCCAAACCCAGACG
 GAAGGGAAACTCGGAGTTTACCGGAATCATAGGAAGGGTTGGGTCTGC

 4660 4670 4680 4690 4700
 GCCCTGTCAGTTCAATTGACTTGGCTAATTGCCAGTGTAGGCCTATGCAA
 CGGGACAGTCAAGTAACGAACCGATTAAACGGTCACATCCGGATACGTT

 4710 4720 4730 4740 4750
 ATTAAGGTAGAACGCACTCCTTAGCGCTCGTTGACTATTCAACTTT
 TAATTCCATCTTGCCTGAGGAATCGCGAGCACTGATAAGTAGTTGAAAA

 4760 4770 4780 4790 4800
 CCTTTAGAAAAGATATTGGTATAAGCACTTCTTAAAAACCATATTCCA
 GGAAAATCTTTCTATAACCATAACCGTGAAGAATTGGTATAAGGT

 4810 4820
 CTCTGGGTGTATTAAATCTAATTTCctcgag
 GAGACCCACATAAAATTAGATTAAAGgagctc

19/27

Figure 4

taataatgga tccccacc**AT** **GGGCATCGAG** CAGGACGGCC TGCACGCCGG CAGCCCCGCC
GCCTGGGTGG AGAGACTGTT CGGCTACGAC TGGGCCAGC AGACCATCGG CTGCAGCGAC
GCCGCCGTGT TCAGACTGAG CGCCCAAGGGC AGACCCGTGC TGTTCTGAA GACCGACCTG
AGCGGGCGCCC TGAACGAGCT GCAGGACGAG GCCGCCAGAC TGAGCTGGCT GGCCACCACC
GGCGTGCCT GCGCCGCCGT GCTGGACGTG GTGACCGAGG CCGGCAGAGA CTGGCTGCTG
CTGGCGAGG TGCCCGGCCA GGACCTGCTG AGCAGCCACC TGGCCCCCGC CGAGAAGGTG
AGCATCATGG CCGACGCCAT GAGAAGACTG CACACCCTGG ACCCCGCCAC CTGCCCTTC
GACCACCAAGG CCAAGCACAG AATCGAGAGA GCCAGAACCA GAATGGAGGC CGGCCTGGTG
GACCAGGACG ACCTGGACGA GGAGCACCAAG GGCCTGGCCC CCGCCGAGCT GTTCGCCAGA
CTGAAGGCCA GAATGCCCGA CGGCGAGGAC CTGGTGGTGA CCCACGGCGA CGCCTGCCCTG
CCCAACATCA TGGTGGAGAA CGGCAGATT AGCGGCTTCA TCGACTGCGG CAGACTGGC
GTGGCCGACA GATACCAGGA CATGCCCTG GCCACCAGAG ACATGCCGA GGAGCTGGC
GGCGAGTGGG CCGACAGATT CCTGGTGCTG TACGGCATCG CCGCCCCCGA CAGCCAGAGA
ATCGCCTTCT ACAGACTGCT GGACGAGTTC TTCTGAATCT TGCAGCTGGT GGGGTTCTGG
*
ATATCCACTA CTTAGccgaa ttctatata

20/27

Figure 5

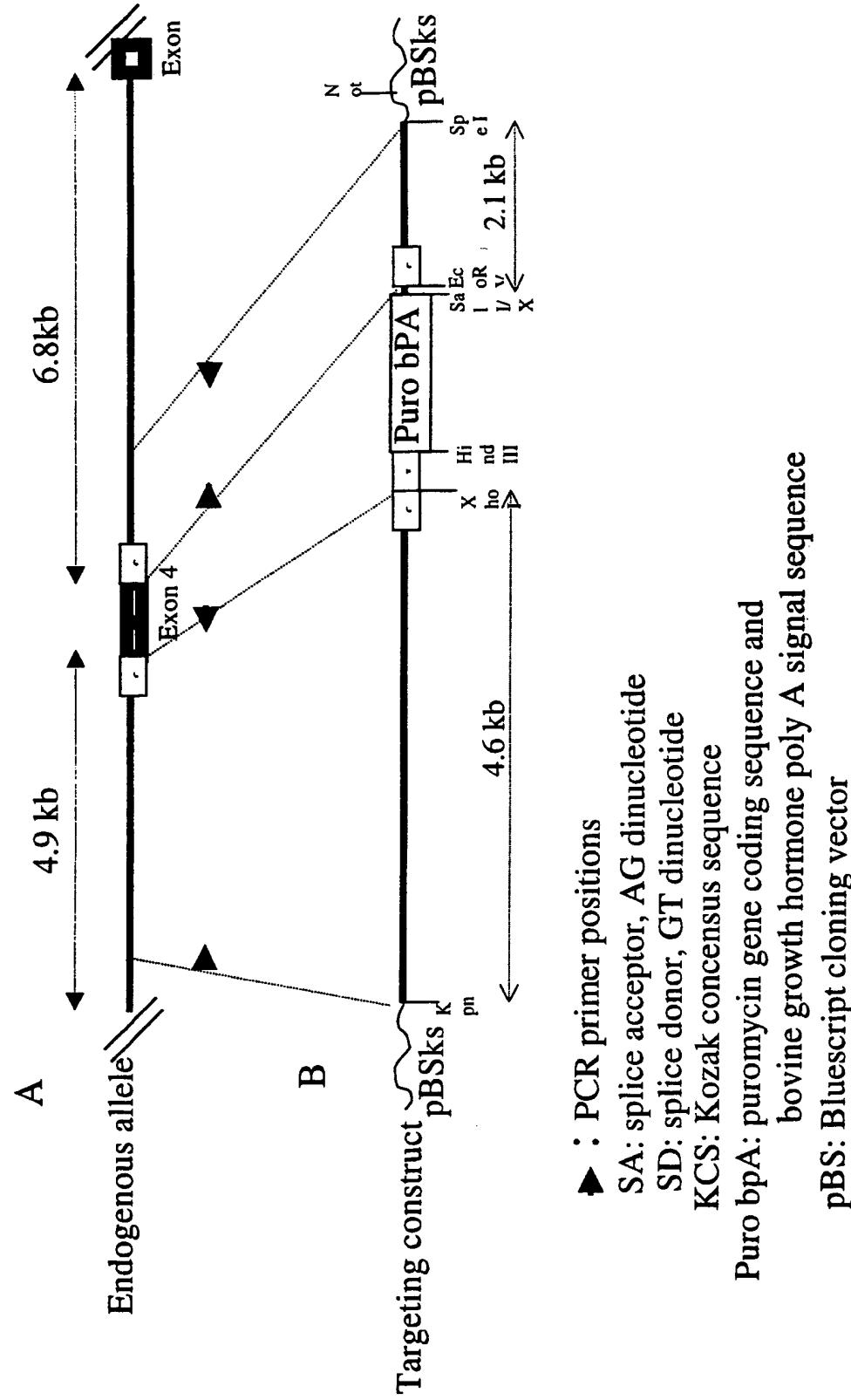
```

1 ctgaagctta ccatgaccga gtacaagccc acggtgccgc tcgccacccg cgacgacgtc
61 ccccgcccg tacgcacccct cgccgcccgcg ttgcggact accccgcac ggcacacacc
121 gtcgaccccg accgcccacat cgagcgggtc accgagctgc aagaactctt cctcacgcgc
181 gtcgggctcg acatcggcaa ggtgtgggtc gcgacgacg gcggccgcgg ggcggcttgg
241 accacgcccgg agagcgtcga aegccccccg gtgtcgccg agatcgccg ggcacatggcc
301 gagttgagcg gttcccgct ggccgcgcag caacagatgg aaggcctctt ggccgcgcac
361 cggcccaagg agcccgctg gtccctggcc accgtcgccg tctcgcccg ccaccaggggc
421 aagggtctgg gcagcgcgtc cgtgtcccccc ggagtggagg cggccgagcg cgccgggggtg
481 cccgccttcc tggagacctc cgcgcggccgc aaccctccct tctacgacg gtcggcttc
541 accgtcaccg cgcacgtcga ggtggcccgaa ggaccgcgc a cctgggtgcat gacccgcgaag
601 cccggtgcct gaccccccc ccacgaccccg cagcgcggcga ccggaaaggag cgcacgaccc
661 catgcatcga tgatctagag ctgggtgatc agcctcgact gtgccttcta gttgccagcc
721 atctgttgtt tgccccctcc cctgacccctt cttgaccctg gaaggtgcca cttccactgt
781 ccttcctaa taaaatgagg aaattgcac c gcattgtctg agtaggtgtc attctattct
841 ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata gcaggcatgc
901 tggggatgcg gtgggcctta tggctctga ggccggaaaga accagcctcg a

```

21/27

Figure 6



22/27

Figure 7

1 gtcgactcta ggcctcaactg gcctaatacg actcaactata gggagctcg a gatcaatta
 61 gaggtccacc atcccttcc tgaatgccta aggccagata tggttgaatt tagaatttt
 121 caaatgcaga atatttatcc tatattatgt aacgccccca gtgcagcaac agcacataac
 181 aatgtacate aatatttatg caaagaatg tttaaacagt ctcactaagt gataaagatt
 241 ataaatagcc tcgtgtcagg gcctggatgc caactgaatt ataaacaggc ttttggttt
 301 cagagcttgg agttggatga gggtctgaga aactgctcca tgttcagggt tacccagtc
 361 gtgggtgtct ccagacccc cctccttccc aagctcttc accacccaca cttctctggg
 421 agtgaagaca acggcagaga ggcatggca cagtggccac agtctccacc ccgatctgtc
 481 tgctcccaa cccaggcett tcctcgact cagtcta gctgttgatg taggatcaa
 541 gtggctttt ccagcatctg ggccgagctg catgtagccc cgtcatttc gtactttgc
 601 cctggggccc gggctgttt tgcaggacc tgaggtgtc acaggaaaga agctccatct
 661 ccccatcttc tcaccatctc tggAACACCA cctatcatga ttgtatctga aagggtggcga
 721 ttgaatcaga ggttccaaa cagagctac tttccaagca agaaggaata gagtgacatg
 781 gctgataatc ccatacttc tcttcttta ctggatttca caacagaggt gatggagcac
 841 ctgagatcta agcctggagt cacccatgaa ccctctctgc aaatatttgg agaataaacc
 901 gtcccctgaa aggacacatc tcagtgcatt tctcatttca ttcacacatc ttttttttt
 961 ttttttttt tttttttgg gcttttgc atttcttggg ccagtcctgc ggcatatgga
 1021 ggtcccccagg ctaagggtct aattggagcc gtagctgcag gcctacgcca gagccagagc
 1081 cacacgggat ctgagccg tctcaac acaccacac tcacggcaac gccggatcct
 1141 taagccactg agcaaggcca gggatggaa ccacaac atgtttccta gtcagat
 1201 ttaaccacag agccacaac ggaactccc cacattattt attgacggcc ttctctgtc
 1261 tctgtggggc actgggaatt cagggtgt caagaagtc tccctcctgc cctcaggaag
 1321 ctcaaaccac tcattat ttgacggct tctctgtc ctgtgggyca ctggaaattc
 1381 aggggtgacg aagaagtcat ccctctgc ctcagga tcaaaaca aggtagagga
 1441 ggcagagcaa aatgcagg ttatccgg agccgactc caggcgtat tgtacagcaa
 1501 aggaatagag ggatggggc cggaggag aaaagggtt cagccgtgt cagggtgggg
 1561 gtggaaatg gcttca aaa ggcagtgaca ttggctcca ggtgtccact ttctctgtc
 1621 tgctaccttc ttgtcctc tttctgg ctcctctatc ctacctctaa agcttca
 1681 acatcctc ttccttct ctctctgg tctctctgg gtaatcaat tggttcc
 1741 cacgtcagat ccgatacgtt cctggtcca tgaacaactt ctccgattgc atggtctg
 1801 tacatcttc tgatgaactt tagacttga tgtccactt tctccctgtc cccttttag
 1861 tattcgcaca ctccccgaca ttcacacgtc caaaaggaa ttcatgatta ttatcctca
 1921 agctctgtcc tcctccagcc catctgagaa aatactacaa ccccctgt taagcagaaa
 1981 tcttgggtct tccttgc atctctgata acaaaattac caaccacgtc ctatcaattc
 2041 tctctca aaatata atata tttaat ttccctgt acagcatggg
 2101 gatcaagtt ttcttacat tatatttcc ccccacc tgttccgtt caatatgat
 2161 atctagacat agttctca gtactc aggatctt tgttaatata agttgtat
 2221 gataacccc agctccc cat ccctccact ccctcc cctgtcgggc agccaca
 2281 ctattctca agtc catgtat tttcttct gtggagatgg tcatttgc tggatatt
 2341 attccagtt taagtgtat catatgtat ttgtca aatat tttttttt
 2401 tcttttgc ttttgtt ttttgtt tgttgc tgttgc tgttgc
 2461 tacttggcc gctccc cg catatggagg ttcc cagg aggatgtaa tggagctgt
 2521 agccacccggc ctac gcca gcccacagca cg gggatcc gagccgc tc
 2581 accacag acgg caac gc tggatc ttt ccactg caagggc agg gacc
 2641 gcaac ctcat ggtt cctgt cg gatt cg ttt ccactg ccac
 2701 gtatattt tttca agcc ccctt gagc cagg ccact cctt tat gtcat
 2761 cggtctgc ttgt ccctt ctc catt tc cacact catgc acc ccagat gt
 2821 gtgaagtt ggat gggt tctcc atgt ttt gggatt ccac ctact
 2881 ttccat acat ttcc agggt tc actgc ttt cccc ctgact gcct agc
 2941 actcaaggg gac gtgt ca gtact ttt tctgt cagg tttt gtttt
 3001 tttcaaa aca cg gac gtc ttt tctgt ctg ctgt ttt ccaaa
 3061 tttcata ccc tttcac ggc ccttgc tat ccat ccat cc ttgg
 3121 gctctaa agt cat ccca aatt gaagg tgact aaaga tttccc aga agg
 3181 gagttccaa ctact acat gacttgc aaat gtttgt gtcttc catta aat gaaaaa

23/27

3241 aaaaaactgt aacaagatata gaaaaatac agaaaggaaa taataagact agaaaagtc
 3301 aatataatgt gaagggtttg catcaaacac taaaataaac tagtacagat gttaaaagac
 3361 taaattatata agttgaagga tagtgtgaa gatgtaaact atgacatcta aaacacaaa
 3421 tggggcggtt cccgtcacgg cacagtggaa acgaatccga cttaggaacca tgagggttgc
 3481 ggttcaattc ctggccctgc tcagtgggtt aaggatccgg tggggcggtg agctgtgg
 3541 tagtagccat tgagggttt gatcccgcgt tgctgtggct ctgggttgcgg
 3601 cagctccgat tcgaccctta gcctggaaac ctccatatgc cgcgggagcg gccccttaaa
 3661 aagacaaaag accaaaaaaa aaaaaaaaaca aaaaacccac aaaatgttgg gaatcagtc
 3721 tctactatgt ttatgttatt gtcaagttt cctttatgt ctgttaatat ttgcgttcta
 3781 gatgttaggt ctctgatata gatgtgcata atgttaacca atgttatgtc ttccctctgg
 3841 attgatccct ttgttattat gtaatgcctt acatttatctt ttgttacatt ctttgg
 3901 gagtattgtc gatatgtggc tagtgcac acattttctt tccttccat ttacaataaa
 3961 tatcttcta tctccacca aattaaagta ctccgcaacc tggttacccaa cccagcatcc
 4021 ctcccttcaactacaat ttcatgcagc gatcaagaaa tagaatgtac cgactgttt
 4081 ccattgggtt ggcattggaa aagtgggtgg aagtgcaga gcttagatta taaaggccag
 4141 ggtgagagtt cccattgtgg tcagctgaaa tgaatctgac tagcatccat gggcac
 4201 gtttgatccc tggcctcaat cagtggtt aaggatctggc gttgtgtcc gtgagttgt
 4261 gtgttagttcg cagacaaggc gtggacttag tggactgtg gctgtggcat aggctagtg
 4321 ctacagctct gattcgaccc ctggctggg aatctctata tgctgtgagt gtggccctaa
 4381 aatttaaatg aaatttaata aaggaccagg gtatattttt ctttgggat aaggtacata
 4441 gtcagtatata caaggacagt agacccatgg aacggatgtc tccctctatgtc tggatgc
 4501 ggtggggcat ctgggtggg ggcggctgya gccccttaggg accattaact aaaccgtca
 4561 ctctccacca ttcgggtgg ccttgggatc agtcaggatc cttccctttt gggctcaaa
 4621 atggccttag ttccttccc aaccaggacg gcccctgtc ttcattgact tggctaatt
 4681 gccaggatgt gctatgca attaaggat aacgcactcc ttacggctcg ttgactatcc
 4741 atcaactttt ctttttagaa aagatattgg tataaggact tcttaaaaaa ccatttcc
 4801 ctctgggtgtt attaatcta atttccctt cttccctt tttccaggA G 3'end of

INTRON 3 -----puromycin bpolyA----- beginning of intron 4

4938 GTa attatgaaac atgatgaaat gatgtttagt aaagtctc
 4981 ctaatctcct agttatcagc caagtcacca gcttgcatta aaagtaggat tcactgacac
 5041 cgtaaagaaa gcattccaga gagttgcgt tgggttcgt gggcagcaaa cccaaatagg
 5101 atccaagagg aggtgggtt gatccctggc cttgttgcgtt ggcttaagga tccggcattt
 5161 ccgtgaccc tgggttaggt tgcagatgca gtcggatct ggcattgtcg tggctgtgc
 5221 gtaggctgtt ggcttcagct ccagtttgc accttgcgtt ggaacttcca tatcccac
 5281 ttggggccct aaaaatcaaa gaaagaaaaga aaatatttca cccttcctgt atccctgagc
 5341 ccttaaatac cgtctttaaa gtcatttagt cttcaagttc cttccagctta attaatttac
 5401 ttccttcctg ccatttttgc attgttctgtt tttttataacc ttcgttgc tgggttaggt
 5461 agagccagaa ataataaggt catgttaaga ccaagatata atattaaattt atttatatga
 5521 ccagatatacg aagtttaccc gagaactttc agacaggaat tccatgagaa atacaccctg
 5581 atttttgcaaa tccctaaata ttttcaggt taaaaggaac aactcaagttt gttgactttt
 5641 gctgcaaaac acaactgagtc gtcgggtt cattttgtcc tggctaaact tttgggtgtt
 5701 ttgtttttttttttaa ctctggaaag caaaatgaat taaacatttc ttagtttca
 5761 aattcatcg tggatttccacc ccaaaatattt gagctgttc tttgttttggaaactac
 5821 tgccttggag atttccagctg gagacgttc tgacagaaaag aatgttgc aacgcgttac
 5881 aaaaatgcattt gatggctttt acttaagagg cattgttacc gcttggcctt tctttcaaaa
 5941 agggccaccc tcaacttggc ctggggcat tccctgtcg gtgcagcggaa aatgaatct
 6001 gacttaggaac cccgggttggg tgggttcaat ccctggcctt gtcgttgc ttaaggatcg
 6061 ggtgttgaag taagctgtgg tggatgttgc agacgcgtt tggatctgg tgggttgc
 6121 ctttgggtgtt ggcggcagc tacagcttca cttggaccctt tagtctggaa accttttag
 6181 tggggccctta aaggaaaaaa agacaacaaa caaacaacaaa accaaaaaaac aacttggc
 6241 ggagagctat gtcatcacca ttgtatattt gatgggttagt gttttagtagt cccctca
 6301 tcaggatgtt ggcctggatt aacatttagaa tggatcttttattt gatgggttagt gttttagtagt
 6361 agcaggacca ttttggccac ttagaaagga actgcattt caggccatc agttagaagga
 6421 ggattctcta gggaggatctc tcttagctca ggggggttcaaa gaattcagtc ttgtccctac
 6481 agcagcttagt gtcactgtca tggcttggct ttgttccctg gcccaggaaat ttctgcattc

24/27

541 tcgggtcga gccaaaaaaaaaaaaaaa aaaaaaaaaaaaaa aaggaggagg nggattccctt agaataagaa
6601 gctatcattt ctttgatgc ttcatagatc taaccacttc tggAACAGTT atccctctc
6661 attctgaaga actcattttt agaaaaacaa gacgagctag agagtgaaca aatggctac
6721 aaaccaggcc ttgcatttgg agggaaactgt ggtacttcctt ctgaagaaaaa gatgacagcgg
6781 ttggatgcag agaccctggg gctcccttag gtacttgagg actgaggaga tatttcagt
6841 ggaggctgga gctagctgc ctggggctgg tccttgcca ccacttcctt cctctgtgac
6901 ttggggcaag ttccctatc tttaaaaatg gggatgatag tagtacctgc ttcatagggt
6961 tgggatataa aataagttt gaaataagca ctaaggcga cgtacttagt aagcgtggc
7021 tggccatcacc accaccacta tcaccatctg tccggaggc agcataggac aggagattt
7081 tggcaataag aaggaagagt tctaggagtt cccgtgtgg tgcaggggaa atgaatccaa
7141 ctaggaacta ggagggttcg gttcaatcc cgcgcctcgc tcagtggtt aaggatccag
7201 tggccatg agctgtgt tagattgcag acatgctag gatctggagt tgctatggct
7261 gtgggttaag ctggcagctg tagctcgat tctaccctt gcctggaaat ttccgtatgc
7321 cacagggttt gcccataaaa gaaaaaaagaa aaagaaaaag aaaaaattct aggggctgaa
7381 agaatctaac agaagagcaa gttccccatg gggttctga cctgagttga gatgcttgc
7441 taggcaacct tcaagctctg aactcttgc ttttttgaat tgcagccaga gttgtacttc
7501 catattttgg gtacttcaca aaattaaaaac acagaagcca aaggcccaga agtgcattt
7561 ggtgctggcc tccctataaag aggggtgtt tgcagtgctg ggcacactct ctcttcacag
7621 taactggagc agattctggc tgctcttcag ggcgttagt tggcacccag actgcagcca
7681 catcattctt caatgtgagg aatcttattt aacatgtca aaaaaaaatgggtttaa aaggcaggag
7741 attctttgcc accttgtaa ttggtctgag gtgagctgag ggcactaacc tttagacaggt
7801 gggtagcaact gtagctaaag aggattacag gagttctgt tggcttag tggtaacaaa
7861 tccaaacttagt atccatgagg attcaggttc gatccctggc ctcgtcagt gggtcaggt
7921 tccgggtttt ctgtgctgt ggtgtaggct ggcagttca ttttattttac ccctagcctg
7981 ggaacttcca tggctgttag gtaaggcccct tgaaaaaaa aaaaaagaga ttcaaaaata
8041 actccatcaa acacatacag ctgtttaaga atgtcatcca ggacagcatt tggtaaaagg
8101 ctagatgaaa aaaaaaaaaaaa aaaaacttag aatttttattt atttattttt tctttttagg
8161 gcccagaccc tggccatgg aaatgcctgg gctagggtg gaatcagage tgcttacacc
8221 acagccatacg ccacccaga tccaaagcccc gtcgtgacc tacaccacag ctcatggcaa
8281 acactggatc cttaaatccac tgagtgaggc caggaattga acccacattt tcatgatgc
8341 tagttgggtt cttaaagccac tgagccacaa gcttagaatt ttagaggtgg aagaacttt
8401 aagagctata ataaagtaat gatgtgtatg gtgattttga tggtagcggc tactagttat
8461 tgagtgtttt cttgtccag gaactccact gttcattttcc tccctgtttt aaaaacagccc
8521 tggaaaggctca gtgttagtcc acatttctag atgaggaata ctgagttcc acaatattaa
8581 atgtgaacgt tcaagtcac atttttagga agatttagt ccagggtctgt ctgacttggg
8641 taacctgggtt aacccttcctt ttagtcaagg gttccattgt tcaggcgatg aaatggagac
8701 ccagtaggtt aaatgactta acagtgaact tatgtccaaac ttcttaatttag aactcagatc
8761 ttctgattca tcatctgggg ctccttcgtt agctgtgtt tgatgccaaa tgctgcgagg
8821 ggtacagtgt gccgtcaagg agaattccctt ccctcaaggg gttatgtctgt agatggagca
8881 ggcagaggtt cccatgaaag cccaaacaaca caggctagaa ggaggatgtc agagagagag
8941 agcaaaggaa cgtgagagtt cagggaggc aagattatgt tggcttgg gatggatcta
9001 tggatgtttt ttattttttt gggggggggg tcttttgc acttcttggg ctgctcccgaa
9061 ggcataatggg ggttcccagg cttaggggtctt aattggagcc gcagccacca gcctacacca
9121 gagccacagc aacgcaggat ctgagccgc tctgcacact tcaccacagc tcacggcaac
9181 ncgggatctt taacccactg agcaagggca gggaccgaac ctgcaacccctt atggttccctt
9241 gtcagattcg ttaagactg cggccacgac ggaactccctt catttagaaaaa tattttattgt
9301 gcacctactg tatgcccaggc attgtgtctgt gttcatacca aagaaggctc agaagagatg
9361 gcatccggc tggccatgg aggatgatataa tgggtttttt ggcgtacact tcagggtgg
9421 tggatgtgtt acctgagggtt tggatgtttt ctggggaaagg aggggtgagat ggggggg
9481 gaggggttttac taaaaagatg ggacgagggtt gcaaatccaa atctataaaat tggatggctgg
9541 agtgcctcgc aggagggtgg ggcttcgtt gtcgtgggtt cagggccctt tcccttcctt
9601 ctggccctt tccctccccc ctcttgcattt atctgaagtc agattccca ggttcaaataa
9661 ctgtttctt ccttagcagt atgacccctt gcaaaaataat ttattgcctt tggccctctg
9721 aggaggaata gaaacctt cattgtactgt tatttagatt taatgagctt atacatgtca
9781 tggctttaga aaggccccca gccaactattt agtattatgt aatattatca gatcaataga
9841 cagatttaga aacaaggcactt ttagagctg ggtccatggg tactgagctt agagggggaaa
9901 ccataggtgg taggaaggca tggatgtttcat tccctaccagg agatgtggac tcccaactgg

26/27

9961 ggcagaaggc agagggagga gatcggggct ttggcagaat ctcaaacaaa tattagtgg
10021 tagtggttt ttgtttctgt tttaagagat gagggcaggc gtttccgatg tggcgcagtg
10081 gaaacgaatc tgactatgtat ccattgaggat gcagggttca tccctggact cactcagtga
10141 gtttaaggatc cggcattgcc gtgagttgtg gtgtnggtca cagacacacgc tcagatctgg
10201 cattgctgtg gctgtgggt aggctagcag ctgtagctcc aattcaaccc ctacgctggg
10261 aacttccaca tgccgcaggc gcaacccaa aagataaatg aataaataaa taaatatgcg
10321 accttccttt ctggggcct ttgcattgtt ttctctctgt taggcacact cttgctaatac
10381 cctcttcaact gggcctccta agtacccctc agaactcagc taaaacatca tcccctcccc
10441 tggggagcct tcgaggctt cctgttaagt gctcctatgc tttcttggag ttttgaagtc
10501 ctataatgat gtgttatca aaatagggtc caccctccct gccagcttct ctacaccaca
10561 gacacatggt gtctgttca gtcaacactg tatgtctggc acttgacatg taacgcatgc
10621 tcagcaggtt tttgttgaat gaatggaggc ggtctgttag agtgcgtcata tatttactga
10681 tcccttcttg taggatggtc tcactgctt ttttagctt agaagtaccc ttttttttt
10741 ttttttaat ggccacaccc atggcatata gaaattccac gaaggaaagga agaaagaaaag
10801 aaagaaaagaa ggaaatttctt gggtcaggga ttgaatccaa gccacaggtg caacctgagc
10861 tgcagggtt gcaacaccac atcccttaac ccactgtgt gggccaggga tcatacctgt
10921 gcatgtacag cgacccaaaggc cacggcagtc agattcttt tcttcttctt tttcttctt
10981 ttcttttctt tttttttttt tttttttttt ggcttttgc ctttcttagg tgcggcatat
11041 ggaggttccc aggcttagtg tcgaatcaga gctgttagacg ccggctaaa ccacagccac
11101 agcaacacag gatccaagcc ttgtctgtca cctacaccac agctcacggc aacgctggat
11161 ccttaacccg ctgagcgagg ccagggattt aacccgcaac ctcatggttc ttagttggat
11221 tcgttaacca ctgagccatg atggaaactc ctgcagtcag attcttaacc cactatgcca
11281 cagcaggaac tcctagaagt gccctttgag gctactctgt agacagctt gagccagcga
11341 ggcaagaccc ttttttctgg aggaagataa atcctgggtt agggatgggt gggctgtgg
11401 ctccctggga cccatctctg gagcctctt ccctcagcaa agccacccctt gacaataaga
11461 gctgccatct atttttttt cttaaacta agatttgata tttttccagag acctccccc
11521 ccaccgtttt atctgagtaa ttctgaaatg acgagagttc cgtgatatca tttttcgat
11581 ctgcagggtt gaaacctggg agtagccaca acccaggctc tcagctcagc ctagggtttc
11641 aatgataatg attgcaaaat agctttctc tgcatccaa gtaacatgat atgaaaaat
11701 ttccatttgc ttttag //

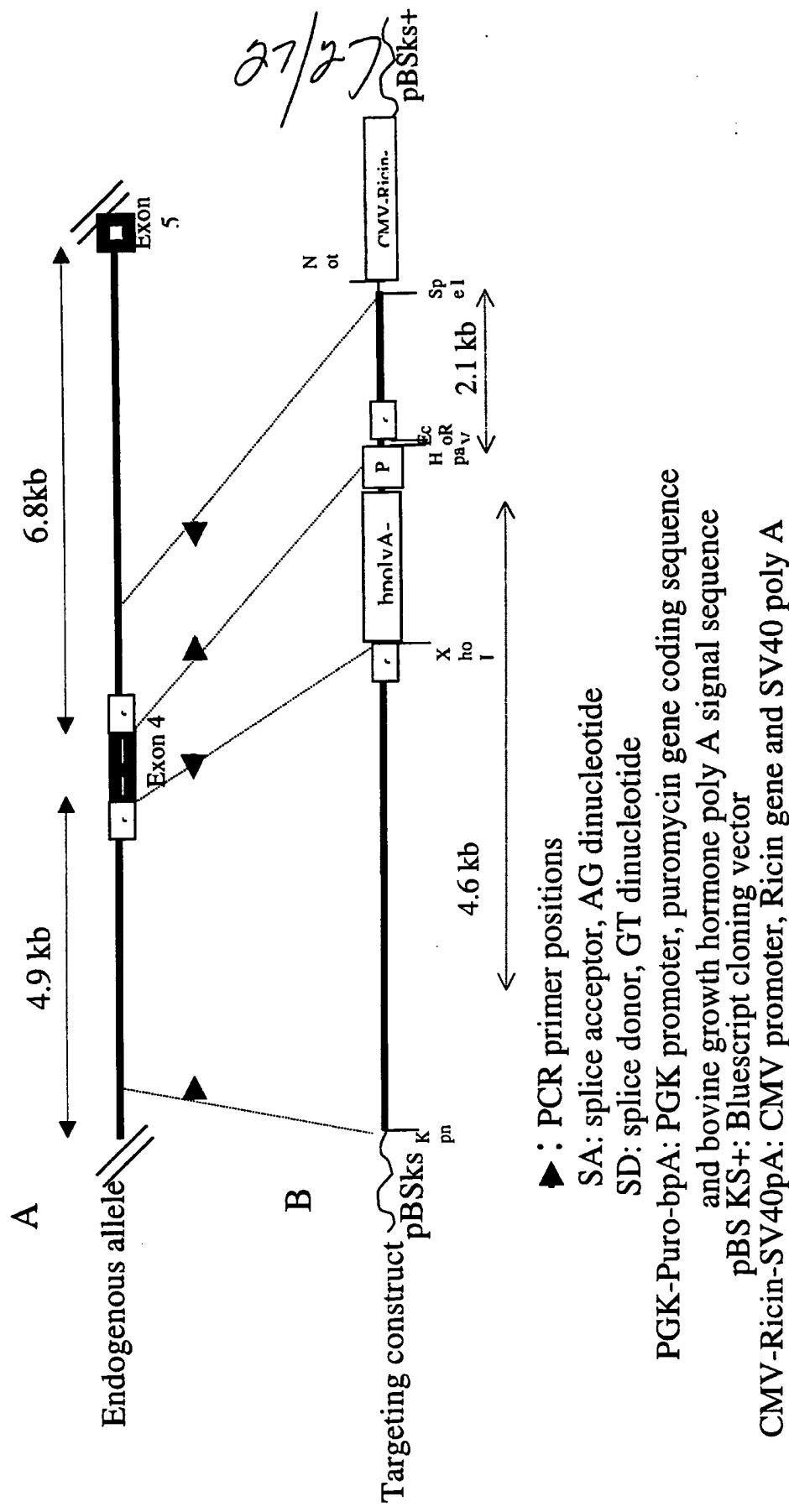
26/27

Figure 8

1 gaattcgcga atccggagtg taacatATGA TATTTCCCAA ACAATACCCA ATCATAAACT
61 TTACCACTGC AGGTGCCACT GTGCCAAAGCT ACACAAACTT CATCAGAGCT GTTCGCGGCC
121 GCTTAACAAAC TGGAGCTGAT GTGAGACATG AAATACCAAGT GTGCCAAAC AGAGTTGGTT
181 TGCCTATAAA CCAACGGTTC ATACTAGTTG AACTCTCAAA TCATGCGAGAG CTTTCTGTTA
241 CATTAGCTCT AGATGTCACC AATGCGTATG TGGTCGGCTA CCGTGCTGGA AATAGCGCAT
301 ATTTCTTCA TCCTGACAAT CAGGAAGACG CCGAAGCAAT CACTCATCTT TTCACTGACG
361 TCCAAAATCG ATATAACATTG GCCTTTGGTG GTAATTACGA TAGACTCGAG CAACTTGCTG
421 GTAATCTGAG AGAAAATATC GAGTTGGAA ACGGTCCGCT AGAGGAGGCT ATCTCAGCGC
481 TTTACTATTA CAGTACTGGT GGTACCCAGC TTCCAACCTCT GGCTCGTTCC TTCATAATAT
541 GCATCCAAAT GATTTCAGAA GCAGCAAGAT TCCAATATAT TGAGGGAGAA ATGCGCACGC
601 GTATTAGGTA CAACCGGAGA TCTGCACCAAG ATCCTAGCGT GATCACACTT GAGAATAGTT
661 GGGGGAGACT GTCGACTGCA ATTCAAGAGT CTAACCAAGG AGCCTTGCT AGTCCAATTC
721 AACTGCAAAG ACGTAATGGT TCGAAATTCA GTGTGTACGA TGTGAGTATA TTAATCCCTA
781 TCATAGCTCT CATGGTGTAT AGATGCGCGC CTCCACCGAG CTCACAGTTT TAGTAAggat
841 ccaagctt

Figure 9

Gal α 1-3Galactosyl Transferase Collision Intron Targeting Construct



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.